A STATISTICAL VIEW OF UNIVERSAL PORTFOLIOS

Cengiz Y. Belentepe

A DISSERTATION

In

Statistics

for the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2005

Supervisor of Dissertation

Graduate Group Chairperson

ACKNOWLEDGEMENT

This dissertation is dedicated

to the memory of my step-father,

Byron D. Paddock,

1946-2004.

Winston Churchill once said of Britain's airmen, "never did so many owe so much to so few." The completion of this dissertation is an opportunity to say "never did so few owe so much to so many," for this dissertation would not have been possible without the support and encouragement of so many people. The greatest contributors were the people who kept me on track with enough encouragement and support to complete the task - my wife, Tomoko Belentepe, my advisor, Adi Wyner, and my statistical family - Larry Brown, Dean Foster, Abba Krieger, Michael Steele and Linda Zhao.

I owe special thanks to my advisor, Adi Wyner, for his outstanding intellectual leadership and guidance over the past several years. I feel very fortunate to have him as an advisor and friend. Without his endless creativity, insightful direction, willingness to question the sacred and uncanny ability to observe the overlooked, many of the results in this dissertation would not have been possible. Thank you for being selfless with your time and ideas, patient with my development and generous with your Starbuck's card. I look forward to a long and rewarding friendship.

Likewise, many thanks go to Michael Steele and Dean Foster for their time and guidance during my time at Wharton. It was Mike's charm, wit and love of gambling that brought me to Wharton. At Wharton, Dean was always the voice of reason for any of our trading strategies, politely reminding us to keep at least one foot on solid intellectual ground. Collectively, their broad interests and research backgrounds provided for many insights and there is nothing like a conversation with either of them to challenge one's way of thinking. I am most grateful for what they taught me on the importance of clarity - a clarity in thinking and a clarity in research.

I am also deeply indebted to Larry Brown and Linda Zhao. Quite frankly, I would never have been able to stay on track without their patience, encouragement and interest in my development. Linda's passion for statistics and her endless compassion for students is truly amazing. Larry is an inspiration to all of us - his compassion for his fellow colleagues, immense enthusiasm for the field, and ability to bring out the best in each of us will never be forgotten.

I am also very grateful for having an exceptional doctoral committee and wish to thank Lawrence Brown, Andreas Buja, Dean Foster, Chris Geczy, and Michael Steele for their careful reading of my dissertation and their perceptive comments.

During the course of my graduate studies, I was faced with many challenges and in every instance, the department and faculty members at Wharton kept me afloat and on track. Your kindness and support never went unnoticed and was always deeply appreciated. Thank you. Special thanks must go to Abba Krieger for providing such

a wonderful research environment at Wharton.

I also enjoyed the camaraderie, friendships and guidance of so many other professors – Andreas Buja, Ed George, Dylan Small, and Jonathan Stroud; fellow students – Lisha Chen, Choongtze Chua, Kartikeya Ghia, David Lewin, Zhihua Qiao, Kenny Shirley, Mouli Venkataramani, Liang Wang, Nicholas Ward, Jonathan Weinberg, Juntian Xu and Peng Zhang; and professional colleagues – Dr. Mark Rzepczynski, Dr. Phil Lipsky, and Dr. Wei Feng at John W. Henry & Co., Inc.

Finally and most importantly, I owe so much to my family without whom I would be lost. My deepest thanks and respect go to my wife Tomoko and our daughter Kapri for their support, encouragement and understanding during these past few years. I could not have done it without their unconditional love and boundless support. Thank you.

While this dissertation marks the end of my formal education, I embark upon a new journey better equipped from my time spent among such wonderful people. Thank you.

ABSTRACT

A STATISTICAL VIEW OF UNIVERSAL PORTFOLIOS

Cengiz Y. Belentepe

Abraham J. Wyner, PhD

The appeal and allure of Cover's universal portfolio is that it presents an implementable investment strategy that in many ways eliminates the regret that one would experience from viewing their investment decisions with the benefit of hindsight. In this thesis, we provide a statistical view of universal portfolios in order to develop a clearer understanding of their performance on actual financial data sequences. By recasting the analysis of a universal portfolio in statistical terms - with a special emphasis on statistical estimation - we are able to resolve a long standing and false perception of a disconnect between information theory and empirical finance. We show that by allowing short sales and leverage the universal portfolio algorithm is approximately equivalent to a sequential Markowitz mean-variance portfolio optimization. If short sales and leverage are not allowed or restricted, then the universal portfolio is approximately equivalent to a constrained sequential optimization. In light of this equivalence, we conclude that universal portfolio construction, while presented as a distributional free procedure, embeds a well known statistical problem - estimation of a multivariate mean and covariance matrix.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Data doesn't come with a model on it's back.

-John Tukey

At the heart of any statistical analysis of financial data lies a critical decision – will a statistical model be *assumed* or not. The importance of this decision should be evident. We need to prevent ourselves from confusing estimation risk with model risk. Assuming a model and showing how small one's estimation error is a high-class sleight of hand.

What this dissertation hopes to achieve is the following. We will take the Universal Portfolio theory, an information theoretical approach to the financial markets developed by Thomas Cover, and we make an unexpected connection to traditional mean-variance portfolio theory as developed by Markowitz [25]. With this connection, we show that at the heart of the Universal Portfolio Theory lies a difficult problem in

statistical estimation – the estimation of a multivariate mean and covariance matrix. Being that we have decided not to assume a model of financial returns, we consider the penalty that one's investment incurs from estimating a multivariate mean, or in our vernacular "μ chasing." This brings us back to where we started and where the field of finance was several decades ago. Without a statistical model for financial returns, a purely statistical approach proves somewhat difficult.

This dissertation is organized as follows. Chapter 2 is motivational and intended to provide the reader with a flavor of what sort of issues we will address in the rest of the dissertation by looking at a few classic gambling problems and accounting for estimation error. Chapter 3 introduces the Universal Portfolio. Chapter 4 establishes the connection between the Universal Portfolio, statistical methodology, and traditional finance. Chapter 5 examines the consequences of estimation error on our portfolio's growth rate. Chapter 6 considers some other Universal Portfolios in the literature. Chapter 7 is to some extent a non-sequitur; it is a stand alone investigation of the expected drawdown in an investment that follows geometric Brownian motion.

Chapter 2

Log-Optimal Investments

The purpose of this chapter is to motivate the examination of the Universal Portfolio. The hope is that through the few simple examples in this chapter , the reader will gain a clearer understanding of the research problems we seek to address in the remainder of this dissertation.

2.1 Motivation

The first task of any gambler or investor is to find a bet or investment opportunity with positive expectation. The second task is to determine how much capital one should bet or invest in this opportunity.

The first task is a perpetual challenge, but the second challenge has useful theoretical support. Kelly [24] answered this question and in his honor, the solution to the problem of optimal bet size allocation is called Kelly betting. Breiman [7] extended

his work by showing that this strategy maximizes the asymptotic growth rate and minimizes the expected time to reach a sufficiently large goal.

In a quick recap of Kelly's result we will restrict ourselves to a gambler making even money bets on the outcome of a coin under the simple premise that more money is better than less and investors are risk adverse. In particular, our gambler is presented with an opportunity to bet as often and as much as he'd like on the successive flips of an independent and identically distributed *biased* coin. In particular, this is a favorable game whereby p, his probability of winning \$1, is greater than $\frac{1}{2}$. Our gambler starts out with a capital of X_0 and wonders whether he should maximize the expected value of his wealth or the expected value of his log wealth.

In financial parlance, we are addressing the issue of determining and maximizing an expected utility function. Any reasonable utility function should be both increasing and concave. Increasing because the more money the better and concave because investors are risk adverse. As it turns out, Kelly betting, or log-optimal betting, is a special case of a class of utility functions - the constant relative risk aversion utility functions, which satisfy these two conditions. While the choice of a utility function may appear subjective at first, in the long run it would seem unreasonable to use anything other than a log utility if more money is truly better. That's because in the long run we are almost surely guaranteed to make more money by maximizing the expectation of log wealth than by maximizing the expectation of any other utility function. Nonetheless, there are well known criticisms of a log utility function, most

notably the one by Merton and Samuelson [27]. We will return to this topic in Chapter 5 and address some of the objections towards log optimal portfolios by showing its connection to the tangency portfolio in a Markowitz [25] mean-variance framework.

2.1.1 Kelly Betting

Bold play will lead our gambler to maximize the expected value of his wealth after n flips of the coin, which we denote as X_n. So if we let

$$\begin{aligned} r_k &= 1 \text{ if the } k^{th} \text{ coin flip is a win} \\ r_k &= -1 \text{ if the } k^{th} \text{ coin flip is a loss} \end{aligned}$$

and b_k be our bet size on the k^{th} bet we see

$$X_k = X_{k-1} + r_k b_k, \quad X_n = X_0 + \sum_{k=1}^{n} r_k b_k$$

$$E[X_n] = X_0 + \sum_{k=1}^{n} E[r_k b_k] = X_0 + (p-q) \sum_{k=1}^{n} b_k.$$

To maximize $E[X_n]$ we maximize b_k on each flip, so bold play is simple: "Bet *all of our wealth* at each flip." If we consider the probability of ruin we quickly see this is a doomed strategy. To see this, consider the random variable A_n defined as

$$A_n = \begin{cases} 1 & \text{Still alive to flip on the } n^{th} \text{ flip} \\ 0 & \text{Bankrupt on the } n^{th} \text{ flip.} \end{cases}$$

Now the probability of these events are given by

$$P(A_n = 1) = p^n = 1 - P(A_n = 0).$$

With $0 \geq p < 1$ we know $\sum_n P(A_n) = \sum_n p^n < \infty$ so the Borel-Cantelli lemma tells us $P([A_n = 1], i.o) = 0$. If we consider the complement of this

$$1 = P(\limsup_{n\to\infty}[A_n = 1]^c) \quad = \quad P(\liminf_{n\to\infty}[A_n = 0]) = 1,$$

we realize from some point on, with probability one, the A_n is identically equal to zero, and we go bust using this strategy almost surely. So, in fact, using a bold betting strategy to maximize expected wealth can be dismissed as quite foolish.

Kelly [24] and Breiman [7] advocate that we maximize the growth rate of our wealth by wagering the same fraction of our wealth on each flip of the coin. Namely, we bet $b_k = fX_{k-1}$ where f is a fixed fraction between 0 and 1. After n flips our wealth is

$$X_n = X_0(1+f)^W(1-f)^L$$

where W = number of wins,L = number of losses and $W + L$ =n. Interestingly enough, since $0 < f < 1$ we know $P(X_n = 0) = 0$, so Kelly betting insures us against ruin in the classical gambler's sense.

Seeking to maximize the expected value of $\log X_n$ for fixed n, we recognize that the maximum of $E[\frac{1}{n}\log X_n]$ with respect to f will lead us to the same f as the maximization of $E[\log X_n]$. As such, we solve for f in

$$\begin{aligned}
\frac{d}{df}E\left[\frac{1}{n}\log X_n\right] &= \frac{d}{df}E\left[\frac{1}{n}\log X_0 + \frac{W}{n}\log(1+f) + \frac{L}{n}\log(1-f)\right] = 0 \\
&= \frac{d}{df}\left(\frac{1}{n}\log X_0 + p\log(1+f) + q\log(1-f)\right) = 0 \\
&= \frac{p}{1+f} - \frac{q}{1-f} = 0
\end{aligned}$$

$$= \frac{p-q-f}{(1+f)(1-f)} = 0.$$

Solving, we find $f_{Kelly} = f_{max} = p - q$; in other words, if we want to maximize our expected log wealth we should always bet our "percentage edge."

Now that we know how to wager our bets in favorable games, how well do we do after N bets? As before, our wealth at time n is

$$X_N = X_0(1+f_{Kelly})^W(1-f_{Kelly})^L$$

so our expected log wealth is

$$\begin{aligned} E[\log X_N] &= \log X_0 + E[W]\log(1+f_{Kelly}) + E[L]\log(1-f_{Kelly}) \\ &= \log X_0 + Np\log 2p + Nq\log 2q \qquad (2.1.1) \end{aligned}$$

which can be rewritten in terms of its entropy $h(p)$ as

$$\begin{aligned} E[\log X_N] &= \log X_0 + N(p\log 2 + p\log p + q\log 2 + q\log q) \\ &= \log X_0 + N\underbrace{[\log 2 - h(p)]}_{\text{growth rate}} \end{aligned}$$

From Jensen's Inequality and the above expression for $E[\log X_n]$ we know

$$\begin{aligned} E[\log X_N] &\leq \log(E[X_N]) \\ X_0 e^{N(p\log 2p + q\log 2q)} &\leq E[X_N]. \end{aligned}$$

Once again, this can be rewritten in terms of entropy as

$$X_0 e^{N(\log 2 - h(p))} \leq E[X_N].$$

This means our expected wealth using Kelly betting is bounded below. Table (2.1) displays these lower bounds on $E[X_N]$ for different values of p and N if we start out with X_0= \$1. I.e. if we could find 250 bets a year with $p = .54$ we would have *at*

	p			
N	.52	.53	.54	.60
125	1.10	1.25	1.49	12.3
250	1.21	1.57	2.22	153
500	1.49	2.46	4.96	23,500

Table 2.1: Lower bounds on expected terminal wealth from Kelly betting with an initial \$ investment.

least a 122% expected return on our initial investment. That's very nice but consider what happens when $p = .60$ – a floor on expected return of 15,200% !

2.1.2 Example: Trading autocorrelation

Consider a time series of returns r_t for a financial asset where our returns are such that $(r_t, r_{t-1}) \sim \mathrm{N}(0, 0, 1, 1, \rho)$ where $\mathrm{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ represents a bivariate normal distribution with means μ_1, μ_2, variances σ_1^2, σ_2^2 and correlation ρ. If we were allowed to place even money bets on the sign of this financial asset's returns we would do quite well with the following trading strategy.

- If yesterday's return r_{t-1} was positive bet that today's return r_t will also be

positive and wager an amount y.

- If yesterday's return r_{t-1} was negative bet that today's return r_t will also be negative and wager an amount y.

Since the probability of winning with this particular trading strategy is

$$P(r_t < 0, r_{t-1} < 0) + P(r_t > 0, r_{t-1} > 0) = \frac{1}{2} + \frac{1}{\pi}\text{arcsin}\rho,$$

we know that Kelly betting would have us wager our "percentage edge." (See Appendix for details.) That is,

$$y = p - q = \frac{2}{\pi}\text{arcsin}\rho.$$

Table (2.2) displays the requisite ρ in order to obtain certain specified odds. It turns

Odds	ρ	Odds	ρ
51-49	.0314	57-43	.2181
52-48	.0628	58-42	.2487
53-47	.0941	59-41	.2790
54-46	.1253	60-40	.3090
55-45	.1564	65-35	.4540
56-44	.1874	70-30	.5878

Table 2.2: Betting odds calculated from the autocorrelation of returns ρ

out that just a small amount of autocorrelation lets us do quite well under a Kelly betting trading strategy. Combining Table (2.2) and Table (2.1) lead to an astonishing

fact: If our returns have an autocorrelation of .094 we can make 250 bets per year with 53-47 favorable odds leading to an annualized expected return of *at least* 57%.

2.1.3 Example: Log-optimal portfolio

For the purposes of this example, allow our investor to have access to unlimited borrowing and also be willing to take an arbitrarily large short position in any security. Under such a setup, consider an investor who seeks to maximize his expected log wealth but faces the task of allocating his wealth among k financial assets where the vector of asset returns, $\mathbf{r}_t$, at time t is modelled as being i.i.d with mean vector μ and second moment matrix $\mathbf{\Sigma}$. Our investor seeks to find a portfolio weight vector $\mathbf{w}$ that solves the following optimization

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} E[\log(1 + \mathbf{w}'\mathbf{r}_t)]. \tag{2.1.2}$$

In view of the Taylor expansion

$$\log(1 + \mathbf{w}'\mathbf{r}_t) = \mathbf{w}'\mathbf{r}_t - \frac{1}{2}\mathbf{w}'\mathbf{r}_t\mathbf{r}'_t\mathbf{w} + o(\mathbf{w}'\mathbf{r}_t\mathbf{r}'_t\mathbf{w}),$$

we have,

$$\begin{aligned}\arg\max_{\mathbf{w}} E[\log(1 + \mathbf{w}'\mathbf{r}_t)] &\approx \arg\max_{\mathbf{w}} \left\{ E[\mathbf{w}'\mathbf{r}_t] - \frac{1}{2}E[\mathbf{w}'\mathbf{r}_t\mathbf{r}'_t\mathbf{w}] \right\} \\ &\approx \arg\max_{\mathbf{w}} \left\{ \mathbf{w}'\mu - \frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w} \right\},\end{aligned}$$

so, as a replacement for problem (2.1.2), we can consider

$$\arg\max_{\mathbf{w}} \left\{ \mathbf{w}'\mu - \frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w} \right\}.$$

Solving this maximization, we see that our optimal weight vector is the solution of

$$\begin{aligned} \frac{d}{d\mathbf{w}}\left\{\mathbf{w}'\mu - \frac{1}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}\right\} &= 0 \\ \mathbf{w}^* &= \boldsymbol{\Sigma}^{-1}\mu. \end{aligned} \tag{2.1.3}$$

2.2 Asymptotic optimality of the log-optimal portfolio

In the previous section, we found the *unconstrained* log-optimal portfolio

$$\begin{aligned} \mathbf{w}^* &= \arg\max_{\mathbf{w}} E[\log(1+\mathbf{w}'\mathbf{r}_t)] \\ &= \boldsymbol{\Sigma}^{-1}\mu \end{aligned}$$

that maximizes an investor's expected log wealth by allocating among k financial assets where the vector of asset returns, $\mathbf{r}_t$, at time t is modelled as being independent and identically distributed with mean μ and second moment $\boldsymbol{\Sigma}$.

In this section, we will prove almost surely that the *constrained* log-optimal investor, whose portfolio weights are restricted to be positive and sum to one, will not do any worse than any investor who uses another investment strategy, who is also subject to these same constraints. But first some preliminaries are in order.

Lemma 2.2.1. *The expected log wealth* $E[\log(1+\mathbf{w}'\mathbf{r}_t)]$ *is concave in* $\mathbf{w}$.

Proof. Concavity of the logarithm means

$$\log\left\{(\lambda\mathbf{w}_1+(1-\lambda)\mathbf{w}_2)'(\mathbf{1}+\mathbf{r}_t)\right\} \geq \lambda\log\left\{\mathbf{w}_1'(\mathbf{1}+\mathbf{r}_t)\right\} + (1-\lambda)\log\left\{\mathbf{w}_2'(\mathbf{1}+\mathbf{r}_t)\right\}.$$

Take expectations and we have that $E[\log(1+\mathbf{w}'\mathbf{r}_t)]$ is concave in $\mathbf{w}$. □

We will make use of the following Kuhn-Tucker conditions to characterize the maximum of this concave function.

Result 2.2.2. *The log-optimal portfolio weights,* $\mathbf{w}^*$, *satisfy the following necessary and sufficient conditions:*

$$E\left(\frac{1+r_{i,t}}{1+\mathbf{w}'^*\mathbf{r_t}}\right) \begin{cases} = 1 & \textit{if } w_i^* > 0, \\ \leq 1 & \textit{if } w_i^* = 0. \end{cases}$$

Proof. See Theorem 15.2.1 of Cover and Thomas [12]. □

One consequence of the Kuhn-Tucker conditions is the following theorem.

Theorem 1. *Let X_T^* represent the wealth achieved by the log-optimal portfolio* $\mathbf{w}^*$ *and let X_T be the wealth resulting from any other causal portfolio. Then*

$$E\left(\frac{X_T}{X_T^*}\right) \leq 1.$$

Proof. Using Result 2.2.2, we see that for a log-optimal portfolio

$$\begin{aligned} E\left(\frac{1+r_{i,t}}{1+\mathbf{w}'^*\mathbf{r_t}}\right) &\leq 1 \\ \sum_i w_i E\left(\frac{1+r_{i,t}}{1+\mathbf{w}'^*\mathbf{r_t}}\right) &\leq \sum_i w_i = 1. \end{aligned}$$

But this last inequality can be rewritten as

$$E\left(\frac{1+\mathbf{w}'\mathbf{r}_t}{1+\mathbf{w}'^*\mathbf{r}_t}\right) = E\left(\frac{X_T}{X_T^*}\right) \leq 1.$$

□

Now we will prove that the wealth of a log-optimal investor exceeds the wealth of any other investor for almost every sequence of returns.

Theorem 2 (Cover and Thomas [12]). *Let the k-dimensional vector of asset returns $\mathbf{r}_t$ at time t be modelled as i.i.d. draws from a distribution with finite variances. Let X_T^* represent the wealth achieved by the log-optimal portfolio $\mathbf{w}^*$ and let X_T be the wealth resulting from any other portfolio. Then*

$$\limsup_{T\to\infty} \frac{1}{T}\log\frac{X_T}{X_T^*} \ \le \ 0 \ a.s.$$

Proof. From Theorem 1, the expectation of $\frac{X_T}{X_T^*}$ is bounded and hence we can make use of Markov's inequality. In particular, we have

$$\begin{aligned} P\left(\frac{X_T}{X_T^*} > m_T\right) &< \frac{1}{m_T} \\ P\left(\frac{1}{T}\log\frac{X_T}{X_T^*} > \frac{1}{T}\log m_T\right) &< \frac{1}{m_T} \end{aligned}$$

Set $m_T = T^2$ and sum over T,

$$\sum_{T=1}^{\infty} P\left(\frac{1}{T}\log\frac{X_T}{X_T^*} > \frac{2\log T}{T}\right) \le \sum_{T=1}^{\infty}\frac{1}{T^2} = \frac{\pi^2}{6} \tag{2.2.1}$$

Now by the Borel-Cantelli lemma,

$$P\left(\frac{1}{T}\log\frac{X_T}{X_T^*} > \frac{2\log T}{T}, i.o.\right) = 0, \tag{2.2.2}$$

which is just another way of stating

$$\limsup_{T\to\infty} \frac{1}{T}\log\frac{X_T}{X_T^*} \ \le \ 0 \text{ with probability 1.}$$

So our log-optimal investor will do as well or better than any other investor to first order in the exponent. □

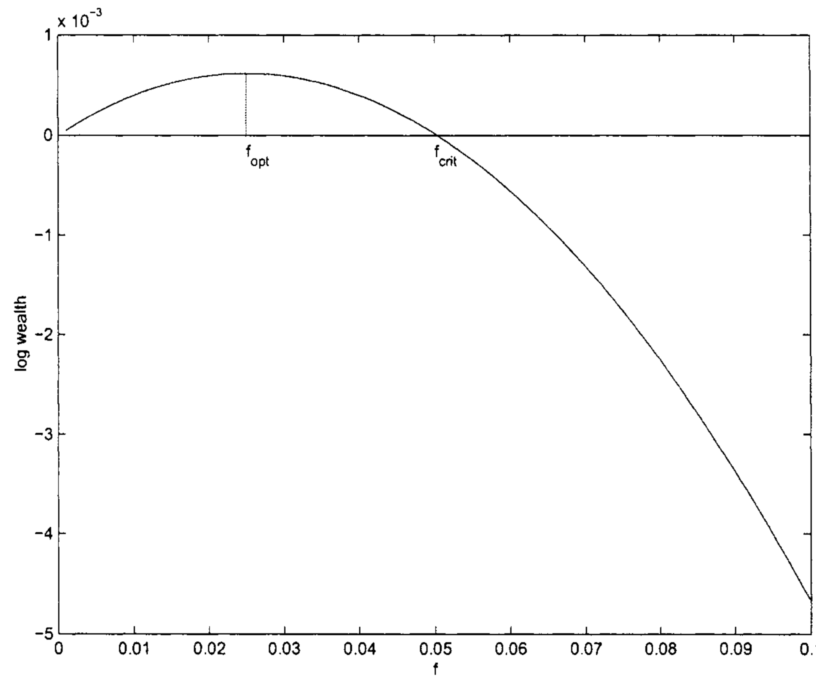

Figure 2.1: Log wealth as a function of fixed fractional f

2.3 All that glitters is not gold

There is an unfortunate consequence of Kelly betting which is apparent from Figure (2.1). If one were to accidently use a fixed percentage f which is beyond f_{crit}, one runs the risk of seeing their expected log wealth approach negative infinity. In other words, one never quite goes bankrupt in the classical sense, but they get arbitrarily close to the poorhouse as time goes on. This critical level, f_{crit}, is the point at which the expected log wealth turns negative. That is

$$p \log(1 + f_{crit}) \quad + \quad q \log(1 - f_{crit}) = 0$$

$$-\frac{p}{q} = \frac{\log(1 - f_{crit})}{\log(1 + f_{crit})}$$

and can be approximated by making use of the following first order expansion

$$\frac{\log(1 - f_{crit})}{\log(1 + f_{crit})} = -1 - f_{crit} + o(f^2),$$

leading to the following approximation of f_{crit}:

$$f_{crit} \approx \tilde{f}_{crit} = \frac{(p-q)}{q}.$$

Unfortunately, the moment we decide to play or invest in any favorable game in the real world, we are faced with the harsh reality that we don't know the true values that p, q or ρ take on and must estimate them from our data. As a consequence, we incur estimation error and must consider the consequences of using $\hat{p}$, $\hat{q}$ and $\hat{\rho}$ instead of p,q and ρ. Unfortunately, the consequences can be quite dire if one calculates the probability of accidently using an $\hat{f}_{Kelly}$ greater than $\tilde{f}_{crit}$.

$$\begin{aligned}
P\,(\text{Negative log wealth}) &= P\left(\hat{f}_{Kelly} \geq \tilde{f}_{crit}\right) = P\left(\hat{p} - \hat{q} \geq \frac{p-q}{q}\right) \\
&= P\left(2\hat{p} - 1 \geq \frac{p-q}{q}\right) = P\left(2\hat{p} \geq \frac{p}{q}\right) \\
&= P\left(\frac{\sum x_i}{n} \geq \frac{p}{2q}\right) = P\left(\sum x_i \geq \frac{np}{2q}\right) \\
&\geq P\left(\sum x_i \geq \left\lceil \frac{np}{2q} \right\rceil\right) = \sum_{k \geq \lceil np/2q \rceil} \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}$$

	p			
N	.52	.53	.54	.60
125	0.327	0.223	0.140	0.002
250	0.243	0.155	0.071	0.000
500	0.173	0.069	0.017	0.000

Table 2.3: Lower bounds on probability of negative log wealth

Table (2.3) displays the non trivial probabilities of incurring the wrath of exceeding $\tilde{f}_{crit}$ for the same values of N and p as in Table (2.1). Even with an 8% favorable edge (54% − 46%) and 125 observations, we still have a 14% probability of using a Kelly wager that exceeds $\tilde{f}_{crit}$.

Similarly, the log-optimal investor who seeks to seeks to use $\mathbf{w}^* = \mathbf{\Sigma}^{-1}\mu$ as his portfolio weight vector will incur estimation error and suffer a similar fate as our Kelly bettor if care is not taken in estimating $\hat{\mu}$ and $\hat{\mathbf{\Sigma}}$. In Chapter 5 we will return to this topic and specifically address the effects of estimation error on one's investment decisions.

2.4 Kelly betting and data compression

Unbeknownst to many, information theory and gambling are related. There exists a strong parallel between the growth rate of one's wealth from repeated bets of a biased coin and the entropy rate of the biased coin.

If in fact we need to estimate $\hat{p}$ and $\hat{q}$ in order to determine our Kelly bets then our expected log wealth, or growth rate, from equation (2.1.1) is

$$\begin{aligned} G(p,q) &= p\log 2\hat{p} + q\log 2\hat{q} \\ &= p\log\left(2\cdot\frac{\hat{p}}{p}\cdot p\right) + q\log\left(2\cdot\frac{\hat{q}}{q}\cdot q\right) \\ &= \log 2 - h\left(p\right) - D\left(p||\hat{p}\right) \end{aligned}$$

where H represents entropy and D represents the Kullback Leibler measure of divergence. In data compression, this divergence is the penalty we incur for using a suboptimal code for p and q. In Chapter 6 we will return to explore the ramifications of this estimation error and how it relates to data compression for the investor seeking a log-optimal portfolio.

2.5 Appendix

Lemma 2.5.1. *Given that* $(r_t, r_{t-1}) \sim N(0,0,1,1,\rho)$ *where* $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ *represents a bivariate normal distribution with means* μ_1,μ_2*, variances* σ_1^2,σ_2^2 *and correlation* ρ*,*

$$P(r_t<0,r_{t-1}<0)+P(r_t>0,r_{t-1}>0) = \frac{1}{2}+\frac{1}{\pi}\arcsin\rho.$$

Proof.

$$\begin{aligned} P(r_t>0,r_{t-1}>0) &= \int_0^\infty\int_0^\infty \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\right\}dxdy \\ &= \frac{1}{2\pi}\int_0^\infty \exp\left\{-\frac{y^2}{2}\right\}\int_0^\infty\frac{1}{\sqrt{1-\rho^2}}\exp\left\{-\frac{(x-\rho y)^2}{2(1-\rho^2)}\right\}dxdy. \end{aligned}$$

Rewritten this is

$$P(r_t > 0, r_{t-1} > 0) \;=\; \frac{1}{2\pi}\int_0^\infty \exp\left\{-\frac{y^2}{2}\right\}\int_{-\frac{\rho y}{\sqrt{1-\rho^2}}}^\infty \exp\left\{\frac{-w^2}{2}\right\} dw dy$$

where

$$w = \frac{x-\rho y}{\sqrt{1-\rho^2}}, \qquad dw = \frac{dx}{\sqrt{1-\rho^2}}.$$

Differentiating with respect to ρ, we have

$$\begin{aligned}
\frac{dP(r_t > 0, r_{t-1} > 0)}{d\rho} &= \frac{1}{2\pi}\int_0^\infty \exp\left\{-\frac{y^2}{2}\right\}\frac{d}{d\rho}\int_{-\frac{\rho y}{\sqrt{1-\rho^2}}}^\infty \exp\left\{-\frac{w^2}{2}\right\} dw dy \\
&= \frac{1}{2\pi}\int_0^\infty \exp\left\{-\frac{y^2}{2}\right\}\left[\exp\left\{-\frac{\rho^2 y^2}{2(1-\rho^2)}\right\}\frac{d}{d\rho}\left(-\frac{\rho y}{\sqrt{1-\rho^2}}\right)\right] dy \\
&= \frac{1}{2\pi}\int_0^\infty \exp\left\{-\frac{y^2}{2}\right\}\left[\exp\left\{-\frac{\rho^2 y^2}{2(1-\rho^2)}\right\}\cdot\frac{-y}{(1-\rho^2)^{\frac{3}{2}}}\right] dy \\
&= \frac{1}{2\pi(1-\rho^2)^{3/2}}\int_0^\infty -y\cdot\exp\left\{-\frac{y^2(1-\rho^2)-\rho^2 y^2}{2(1-\rho^2)}\right\} dy \\
&= \frac{1}{2\pi(1-\rho^2)^{3/2}}\int_0^\infty -y\cdot\exp\left\{-\frac{y^2}{2(1-\rho^2)}\right\} dy. \qquad (2.5.1)
\end{aligned}$$

Substituting

$$v = \frac{-y^2}{2(1-\rho^2)}, \qquad dv = \frac{-y}{(1-\rho^2)}$$

in equation (2.5.1) leads to

$$\begin{aligned}
\frac{dP(r_t > 0, r_{t-1} > 0)}{d\rho} &= \frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^0 e^v dv \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}}.
\end{aligned}$$

Using the above to evaluate $P(r_t > 0, r_{t-1} > 0)$ we have

$$P(r_t > 0, r_{t-1} > 0) \;=\; \int \frac{dP(r_t > 0, r_{t-1} > 0)}{d\rho}$$

$$
\begin{aligned}
&= \frac{1}{2\pi}\int \frac{1}{\sqrt{1-\rho^2}}d\rho \\
&= \frac{1}{2\pi}\text{arcsin}\rho + C.
\end{aligned}
$$

By symmetry arguments, $P(r_t > 0, r_{t-1} > 0) = \frac{1}{4}$ when $\rho = 0$ so C must be equal to $\frac{1}{4}$. Using symmetry once again this means

$$
P(r_t > 0, r_{t-1} > 0) = \frac{1}{4} + \frac{1}{2\pi}\text{arcsin}\rho
$$

and

$$
P(r_t < 0, r_{t-1} < 0) = \frac{1}{4} + \frac{1}{2\pi}\text{arcsin}\rho.
$$

Hence

$$
P(r_t < 0, r_{t-1} < 0) + P(r_t > 0, r_{t-1} > 0) = \frac{1}{2} + \frac{1}{\pi}\text{arcsin}\rho.
$$

□

Chapter 3

Universal Portfolio

3.1 Constant Rebalanced Portfolios

3.1.1 Motivation

The Kelly strategy which maximizes the expected logarithm of wealth in each period has many attractive properties. For an investor to fully appreciate the optimality of this criterion one must consider the behavior of one's terminal wealth when using a constant rebalanced portfolio.

Definition 3.1.1. A constant rebalanced portfolio (CRB) is simply a portfolio whose asset allocation is reset at the end of each time period to the allocation used at the beginning of that period. In essence, a constant rebalanced portfolio attempts to maintain a fixed proportion of wealth in each investment at all times.

More formally, consider an investor who faces the task of allocating his wealth among

k financial assets at the start of trading session t. We denote the vector of asset returns at time t by $\mathbf{r}_t$ and the $k \times 1$ portfolio weight vector as $\mathbf{w}$. The wealth achieved by a constant rebalanced portfolio, starting with an initial wealth of X_0, can be expressed as

$$X_T = X_0 \prod_{t=1}^{T} (1 + \mathbf{w}'\mathbf{r}_t).$$

When should an investor decide to use such a constant rebalanced portfolio? When every day is the same as every other day and independent of the past. So what's best on one day is equally best on every other day. Later, this point will be emphasized in our discussion on universal portfolios. That's because even in the absence of any distributional assumptions on $\mathbf{r}_t$, the use of a constant rebalanced portfolio rests on an assumption that our market returns $\mathbf{r}_t$ be independent and identically distributed. In this case, we see from X_T that the terminal wealth becomes a product of independent and identically distributed terms which after taking logarithms becomes a sum of independent and identically distributed terms. As a consequence, we can state,

$$\frac{1}{T} \log X_T = \frac{1}{T} \log X_0 + E \log(1 + \mathbf{w}'\mathbf{r}).$$

The law of large numbers applies here, and after slight rearrangement shows us the asymptotic growth rate converges to

$$\lim_{T \to \infty} \frac{1}{T} \log \frac{X_T}{X_0} = E \log(1 + \mathbf{w}'\mathbf{r}).$$

Thus the growth rate, or Kelly criterion, is maximized when using portfolio weights $\mathbf{w}$ such that $E \log(1 + \mathbf{w}'\mathbf{r})$ is maximized or equivalently, when $E \log X_T$ is maximized.

3.1.2 Example of how to turn two nags into a thoroughbred

The case for a constant rebalanced portfolio can be made by considering the following example. Imagine there are two stocks which will both eventually go to zero. Could there be a way to make a fortune investing in them ? The answer is yes.

Consider the following two *real world* stocks - Airborne Inc. (Ticker symbol: ABF), listed on the New York Stock Exchange and Hallmark Financial Services, Inc. (Ticker symbol: HAF.EC), listed on the American Stock Exchange. From January 2, 1998 to December 31, 2002, a period of 1256 trading days, each of these stocks went down substantially. In particular, \$1000 invested in Airborne Inc. was worth only \$499 at the end of the period. Likewise, \$1000 invested in Hallmark Financial Services, Inc. was worth only \$566 at the end of the same period. All in all, either one of these investments proved to be rather dismal on their own.

Yet, if we re-balance our portfolio every day to maintain w % of our total wealth in Airborne Inc. and the remainder, $(1 - w)\%$, of our total wealth in Hallmark Financial Services, Inc., then as long as w is between 0.0745 and 0.912 we'll take two nags and turn them into a thoroughbred. Consider Figure (3.1) which shows the terminal wealth of an initial investment of \$1000 for various levels of w. With a daily re-balancing percentage of $w = 50\%$ our initial investment of \$1000 grows to be worth

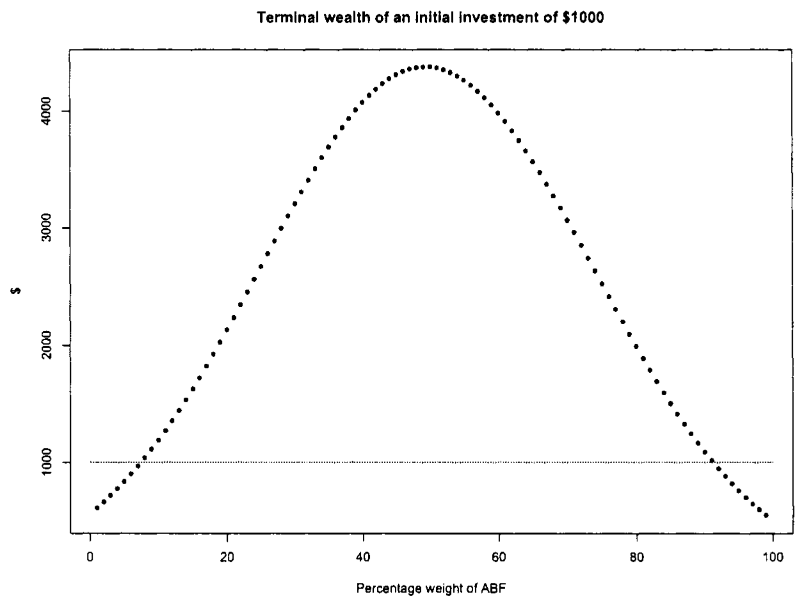

Figure 3.1: Terminal wealth of an initial $1000 investment in Airborne Inc and Hallmark Financial Services, Inc as a function of w from January 2, 1998 to December 31, 2002.

\$4378 by the end of the period - a 34.2% annualized positive return. This is in sharp contrast to a sole investment in HAF.EC, $w = 0\%$, which leads to an annualized loss of 10.7% or to a sole investment in ABF, $w = 100\%$, which leads to an annualized loss of 12.9%.

To further aid one in understanding how these various investments progress through time consider Figure (3.2). Five *separate* investment strategies are displayed in this graph. The first two series are simply the pure investment in one single stock. The last three series display what happens to our wealth as we re-balance everyday to maintain $w\%$ of our wealth in Airborne Inc (ABF) and $(1 - w)\%$ in Hallmark Financial Services, Inc. (HAF.EC) for $w = 25\%$, 50% and 75%.

This is truly a surprising piece of investment advice. Consider the fact that at all times *all of our wealth* is dedicated to a portfolio whose individual stocks each go to zero yet we make money by ensuring at the start of every day we have $w\%$ in ABF and $(1 - w)\%$ in stock HAF.

3.2 Universal Portfolio

3.2.1 Background

The traditional approach of mean-variance portfolio theory, as developed by Markowitz [25] and Sharpe [29], is formulated under a distributional framework. The approach assumes asset returns are governed by a probability distribution and given knowledge

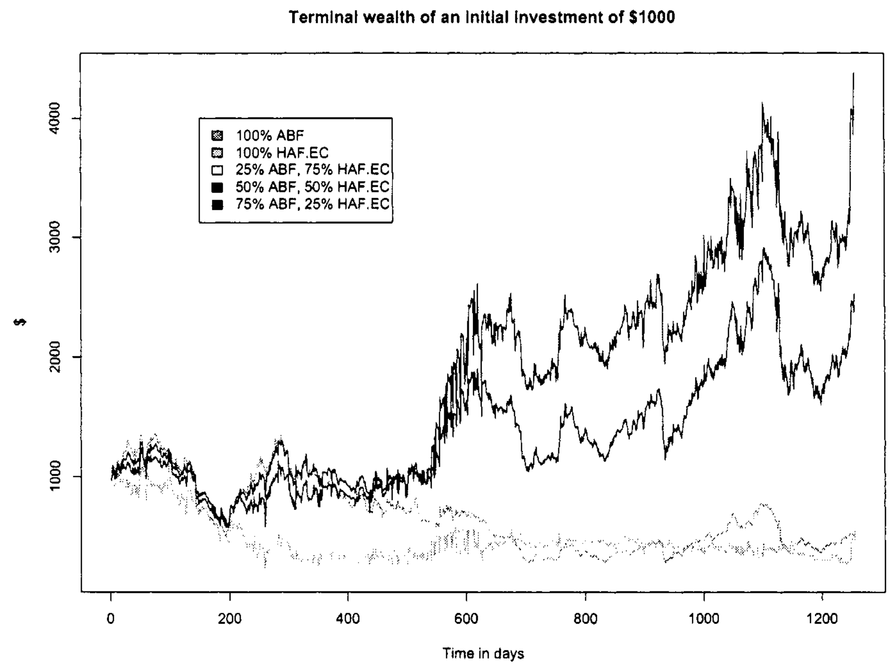

Figure 3.2: Time series of the wealth from an initial investment of $1000 in Airborne Inc. and Hallmark Financial Services, Inc. under various constant rebalanced portfolio weights.

of this distribution, a sequence of investment decisions can be specified to achieve an "optimal" growth rate. In particular, an investor would choose a sequence of portfolio weights that maximizes the first moment of wealth subject to a constraint on the variance. In other words, our investor maximizes expected return for a given level of risk.

Likewise, the theory of rebalanced portfolios as developed by Thorp [33], Cover [8], Markowitz [26], Hakannson [15], Bell and Cover [3] [4], Barron and Cover [2] and Algoet and Cover [1] assumes the underlying distribution for market returns is known.

A significant departure from these model-inspired investment theories is connected to game theory and data compression and arises by assuming the investor has no knowledge of the true distribution underlying the market. Instead, a set of allowable portfolio decisions is defined and the investor's goal is to achieve the same asymptotic growth rate as the best portfolio in this set. Furthermore, the goal is to achieve this uniformly for all possible sequences of market returns.

3.2.2 Notation

Imagine our returns $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t$ are an arbitrary sequence of vectors and instead of maximizing the asymptotic growth rate in an almost sure sense we seek to maximize this growth rate uniformly over all return sequences. To do this, imagine we are allowed to choose any portfolio sequence from a set of such portfolio weight se-

quences W. Obviously, if we had a crystal ball for future market returns $\mathbf{r}_t$, we would just choose the optimal portfolio weight sequence, $\mathbf{w}_1^*, \mathbf{w}_2^*, \ldots, \mathbf{w}_t^*$, from set W that maximizes our actual growth rate $G_t^* \stackrel{\text{def}}{=} \frac{1}{t} \log X_t^*/X_0$. But what we really want is a non-anticipating portfolio weight sequence $\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, \ldots, \hat{\mathbf{w}}_t$ that has a growth rate $\hat{G}_t$ asymptotically equal to G_t^*. Such a portfolio, if it exists, would attain almost the same wealth as if you were given a crystal ball and allowed to act according to any portfolio sequence in W, but without actually having this crystal ball.

Definition 3.2.1. Universal Portfolio. A portfolio weight sequence, $\{\hat{\mathbf{w}}_t\}_{t=1}^T$, is called *universal* with respect to its *target class* W, if its corresponding growth rate $\hat{G}_T$ satisfies

$$\lim_{T\to\infty} \sup_{\{\mathbf{r}_t\}_{t=1}^T} \left(G_T^* - \hat{G}_T\right) = \lim_{T\to\infty} \sup_{\{\mathbf{r}_t\}_{t=1}^T} \frac{1}{t} \log \frac{X_T^*}{\hat{X}_T} \le 0 \tag{3.2.1}$$

$$X_T^* = X_0 \prod_{t=1}^T (1 + \mathbf{w}'^*_t \mathbf{r}_t) \quad , \quad \hat{X}_T = X_0 \prod_{t=1}^T (1 + \hat{\mathbf{w}}'_t \mathbf{r}_t)$$

where G_T^* is the *target growth rate* -- the growth rate of the best hindsight portfolio in W. Similarly, X_T^* and $\hat{X}_T$ are the wealths of the best hindsight and universal portfolio respectively.

Cover [9] named such a sequence $\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, \ldots, \hat{\mathbf{w}}_t$ of portfolios a *universal portfolio* because the above stated optimality is accomplished *universally* over all sequences of market returns.

3.2.3 Prior Work

The goal of achieving X_T^* as specified by equation (3.2.1) is related to the sequential compound Bayes decision problem of Hannan and Robbins [16]. In their study, a player is pitted against nature in a sequence of repeated plays of a game where the player's objective is to develop a sequential strategy that approximates the performance of the best fixed strategy determined in hindsight for any sequence of moves by nature. Hannan and Robbins' [16] methods were applied by Cover and Gluss [10] in the context of investment decisions when the market returns are restricted to a finite set. Cover and Gluss [10] use Blackwell's approach-exclusion theorem to define the portfolio allocation rule

$$\mathbf{w}_{t+1} = \mathbf{w}^*(\hat{\mathbf{p}}_t).$$

Here $\mathbf{w}^*(\hat{\mathbf{p}})$ is the log optimal portfolio weight vector with respect to the probability mass function $\hat{\mathbf{p}}_t$ which is unobserved but known to be a function of the empirical probability mass function $\mathbf{p}_t$ of market returns. Nonetheless, Cover and Gluss prove that this allocation rule tracks the best constant rebalanced portfolio to within an asymptotic factor of $e^{c\sqrt{t}}$. The authors even state that their

> *"sequential portfolio algorithm outperforms an investor who has knowledge of the individual stock prices in the far future."*

These authors fail to appreciate the irony that they have tied this prescient investor's hands by forcing him to commit to one strategy and only one strategy on day one.

Naturally, one wonders if every *target class* W has a universal portfolio? A moment's reflection should convince one that a universal portfolio may not exist for every target class. Having said that, a universal portfolio was first put forth by Cover [9] on a very specific target class – constant rebalanced portfolios. As mentioned earlier, a constant rebalanced portfolio is a sequence of portfolio weights for which one's wealth in each asset is rebalanced at the end of every time period to bring the allocations back to some fixed initial value. The set of all such portfolios will be denoted W_{CRB} and the wealth achieved by CRB $\mathbf{w}$ is as before

$$X_T = X_0 \prod_{t=1}^{T} (1 + \mathbf{w}'\mathbf{r}_t).$$

Definition 3.2.2. The *best constant rebalanced portfolio* is the portfolio $\mathbf{w}^*$ that maximizes wealth X_T in hindsight. X_T^* denotes the wealth achieved by this *best constant rebalanced portfolio* and $G_T^* \stackrel{\text{def}}{=} \frac{1}{T} \log X_T^*/X_0$ denotes it's growth rate.

The first universal portfolio with respect to the target class W_{CRB} where X_T^*, the wealth of the *best constant rebalanced portfolio*, is the target wealth and G_T^*, the growth rate of the *best constant rebalanced portfolio*, is the target growth rate was put forth in Cover [9].

3.2.4 Best Constant Rebalanced Portfolio

There are some notable properties of the target wealth X_T^*.

Proposition 3.2.3. *The best constant rebalanced portfolio exceeds the best stock.*

$$X_T^* \geq \max_{i=1,\dots,k} \prod_{t=1}^{T} (1 + \mathbf{e}_i' \mathbf{r}_t), \tag{3.2.2}$$

where $\mathbf{e}_i$ *is the* i^{th} *basis vector* $(0, 0, \dots, 0, 1, 0, \dots, 0)'$.

Proof. X_T^* is maximized over all constant rebalanced portfolios in the simplex W which includes the portfolio that invests all of its wealth in just one stock. That is, maximization on the right hand side of the inequality is over a subset of the simplex W – the vertices. □

Proposition 3.2.4. *The best constant rebalanced portfolio exceeds any arithmetic index like the Dow Jones Industrial Average (DJIA) or the Standard & Poor's 500 (S&P 500). If* $\alpha_i \geq 0, \sum \alpha_i = 1$, *then*

$$X_T^* \geq \sum_{i=1}^{k} \alpha_i \prod_{t=1}^{T} (1 + \mathbf{e}_i' \mathbf{r}_t).$$

Proof. Starting from inequality (3.2.2) we have

$$X_T^* \geq \max_{i=1,\dots,k} \prod_{t=1}^{T} (1 + \mathbf{e}_i' \mathbf{r}_t) \geq \prod_{t=1}^{T} (1 + \mathbf{e}_i' \mathbf{r}_t)$$

but since $\sum \alpha_i = 1$, we know

$$X_T^* = \sum_{i=1}^{k} \alpha_i X_T^* \geq \sum_{i=1}^{k} \alpha_i \prod_{t=1}^{T} (1 + \mathbf{e}_i' \mathbf{r}_t).$$

□

Proposition 3.2.5. *The best constant rebalanced portfolio exceeds any geometric index.*

$$X_T^* \geq \left\{ \prod_{i=1}^{k} \left(\prod_{t=1}^{T} (1 + \mathbf{e}_i' \mathbf{r}_t) \right) \right\}^{1/k}$$

Proof. Once again, starting from inequality (3.2.2) we have

$$\begin{aligned}
X_T^* \geq \max_{i=1,\ldots,k} \prod_{t=1}^{T}(1+\mathbf{e}_i'\mathbf{r}_t) &\geq \prod_{t=1}^{T}(1+\mathbf{e}_i'\mathbf{r}_t) \\
\prod_{i=1}^{k} X_T^* &\geq \prod_{i=1}^{k}\left(\prod_{t=1}^{T}(1+\mathbf{e}_i'\mathbf{r}_t)\right) \\
X_T^* &\geq \left\{\prod_{i=1}^{k}\left(\prod_{t=1}^{T}(1+\mathbf{e}_i'\mathbf{r}_t)\right)\right\}^{1/k}.
\end{aligned}$$

□

While every investor would be ecstatic to achieve X_t^*, the simple fact of the matter is one can not use the best CRB since it requires knowledge of future returns. However, Cover's universal portfolio construction circumvents this obstacle as follows.

Cover's strategy is to have our portfolio weights specified by

$$\hat{\mathbf{w}}_1 = \left(\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k}\right), \qquad \hat{\mathbf{w}}_{t+1} = \frac{\int_W \mathbf{w} X_t(\mathbf{w}) d\mathbf{w}}{\int_W X_t(\mathbf{w}) d\mathbf{w}} \tag{3.2.3}$$

where $W = \left\{\mathbf{w} \in \mathbf{R}^k : w_i \geq 0, \sum_{i=1}^k w_i = 1\right\}$ and

$$X_t(\mathbf{w}) = \prod_{s=1}^{t}(1+\mathbf{w}'\mathbf{r}_s) \tag{3.2.4}$$

represents the wealth achieved by using $\mathbf{w}$ as our CRB portfolio weight vector and integration in equation (3.2.3) is taken over all constant rebalanced portfolios in W.

The wealth $\hat{X}_t$ achieved by Cover's universal portfolio is given by

$$\hat{X}_T = \prod_{t=1}^{T}(1+\hat{\mathbf{w}}_t'\mathbf{r}_t), \tag{3.2.5}$$

and Cover and Ordentlich [11] prove the following:

Theorem 3. *For a portfolio of k assets, the Universal Portfolio algorithm satisfies*

$$\frac{X_T^*}{\hat{X}_T} \leq (T+1)^{k-1}$$

for every T.

Proof. See Cover and Ordentlich [11]. $\square$

Incidentally, a more intuitive characterization exists for how the universal portfolio works.

Lemma 3.2.6. *The Universal Portfolio can be characterized as the average of wealths from every constant rebalanced portfolio in W*

$$\hat{X}_T = \prod_{t=1}^{T}(1+\hat{\mathbf{w}}_t'\mathbf{r}_t) = \frac{\int X_T(\mathbf{w})d\mathbf{w}}{\int d\mathbf{w}}. \tag{3.2.6}$$

Proof. Substituting the expression for $X_T(\mathbf{w})$ from equation (3.2.4) into equation (3.2.3), we have

$$\begin{aligned}
(1+\hat{\mathbf{w}}_t'\mathbf{r}_t) &= 1+\frac{\int \mathbf{w}'\mathbf{r}_t \prod_{s=1}^{t-1}(1+\mathbf{w}'\mathbf{r}_s)\,d\mathbf{w}}{\int \prod_{s=1}^{t-1}(1+\mathbf{w}'\mathbf{r}_s)\,d\mathbf{w}} \\
&= \frac{\int (1+\mathbf{w}'\mathbf{r}_t)\prod_{s=1}^{t-1}(1+\mathbf{w}'\mathbf{r}_s)\,d\mathbf{w}}{\int \prod_{s=1}^{t-1}(1+\mathbf{w}'\mathbf{r}_s)\,d\mathbf{w}} \\
&= \frac{\int \prod_{s=1}^{t}(1+\mathbf{w}'\mathbf{r}_s)\,d\mathbf{w}}{\int \prod_{s=1}^{t-1}(1+\mathbf{w}'\mathbf{r}_s)\,d\mathbf{w}}.
\end{aligned}$$

Once we substitute in the above, equation (3.2.5) telescopes to

$$\begin{aligned}
\hat{X}_T &= \prod_{t=1}^{T}(1+\hat{\mathbf{w}}_t'\mathbf{r}_t) \\
&= \frac{\int (1+\mathbf{w}'\mathbf{r}_1)\,d\mathbf{w}}{\int d\mathbf{w}} \cdots \frac{\int \prod_{s=1}^{T-1}(1+\mathbf{w}'\mathbf{r}_s)\,d\mathbf{w}}{\int \prod_{s=1}^{T-2}(1+\mathbf{w}'\mathbf{r}_s)\,d\mathbf{w}} \cdot \frac{\int \prod_{s=1}^{T}(1+\mathbf{w}'\mathbf{r}_s)\,d\mathbf{w}}{\int \prod_{s=1}^{T-1}(1+\mathbf{w}'\mathbf{r}_s)\,d\mathbf{w}}
\end{aligned}$$

$$= \frac{\int \prod_{s=1}^{T} (1 + \mathbf{w}'\mathbf{r}_s)\, d\mathbf{w}}{\int d\mathbf{w}}$$

$$= \frac{\int X_T(\mathbf{w}) d\mathbf{w}}{\int d\mathbf{w}}. \tag{3.2.7}$$

□

This is just a formal way of saying the universal portfolio's wealth can be characterized as the average of the wealths from each constant rebalanced portfolio in W. Cover [9] provides an intuitive explanation of what his universal portfolio is actually doing. Cover states

> *The main idea of the portfolio algorithm is quite simple. The idea is to give an amount* $d\mathbf{w}/\int_W d\mathbf{w}$ *to each portfolio manager indexed by rebalancing strategy* $\mathbf{w}$*, let him make* $X_t(\mathbf{w}) = e^{tG(\mathbf{w})}d\mathbf{w}/\int d\mathbf{w}$ *at exponential rate* $G(\mathbf{w})$ *and pool the wealth at the end. Of course, all dividing and repooling is done "on paper" at time* t*, resulting in* $\hat{\mathbf{w}}_t$*. Since the average of exponentials has, under suitable smoothness conditions, the same asymptotic growth rate as the maximum, one achieves almost as much wealth achieved by the best constant rebalanced portfolio.*

Essentially, this is all there is to Cover's theory. By allowing a Riemann sum to approximate the integral in equation (3.2.7), we see that on day one we determine all possible constant rebalanced portfolios in W where we impose some discretization on the continuum of possibilities. Then we divide our initial wealth X_0 equally among these allowable constant rebalanced portfolios and assign a different portfolio manager

to run each unique $\mathbf{w}$ constant rebalanced portfolio. At the end of each time period, each portfolio manager rebalances according to his unique $\mathbf{w}$, and we aggregate all their performances to see how well our *universal portfolio* has done. Cover's universal portfolio is just the limit as the discretization goes to zero, and the proof requires some regularity conditions.

Chapter 4

Statistics, Finance and the Universal Portfolio

At first glance, Cover's theory of a universal portfolio appears unrelated to any existing theories in finance. Upon closer inspection this turns out to be false. In fact, there are direct parallels between Cover's universal portfolio and traditional mean-variance portfolio theory.

Recall that Cover's operational definition of the universal portfolio weights at time $t+1$ is

$$\hat{\mathbf{w}}_{t+1} = \frac{\int \mathbf{w} X_t(\mathbf{w}) d\mathbf{w}}{\int X_t(\mathbf{w}) d\mathbf{w}},$$

which can be rewritten as

$$\hat{\mathbf{w}}_{t+1} = \frac{\int \mathbf{w} e^{tG_t(\mathbf{w})} d\mathbf{w}}{\int e^{tG_t(\mathbf{w})} d\mathbf{w}} \tag{4.0.1}$$

where

$$
\begin{aligned}
G_t(\mathbf{w}) &\stackrel{\text{def}}{=} \frac{1}{t}\log X_t(\mathbf{w}) \\
&= \frac{1}{t}\sum_{s=1}^{t}\log(1+\mathbf{w}'\mathbf{r}_s) \qquad (4.0.2)
\end{aligned}
$$

It follows using the first two terms of a Taylor series expansion, that

$$
\begin{aligned}
G_t(\mathbf{w}) &= \frac{1}{t}\sum_{s=1}^{t}\left(\mathbf{w}'\mathbf{r}_s - \frac{1}{2}\mathbf{w}'\mathbf{r}_s\mathbf{r}_s'\mathbf{w}\right) + o\left(\mathbf{w}'\mathbf{r}_s\mathbf{r}_s'\mathbf{w}\right) \quad \text{as} \quad \mathbf{w}'\mathbf{r}_s \to 0 \\
&= \mathbf{w}'\bar{\mathbf{r}}_t - \frac{1}{2}\mathbf{w}'\bar{\mathbf{\Sigma}}_t\mathbf{w} + o\left(\mathbf{w}'\mathbf{r}_s\mathbf{r}_s'\mathbf{w}\right) \quad \text{as} \quad \mathbf{w}'\mathbf{r}_s \to 0 \qquad (4.0.3)
\end{aligned}
$$

with

$$
\bar{\mathbf{r}}_t \stackrel{\text{def}}{=} \frac{1}{t}\sum_{s=1}^{t}\mathbf{r}_s, \qquad \bar{\mathbf{\Sigma}}_t \stackrel{\text{def}}{=} \frac{1}{t}\sum_{s=1}^{t}\mathbf{r}_s\mathbf{r}_s'.
$$

There are many approaches to justify the Taylor series approximation for $\log(1+\mathbf{w}'\mathbf{r}_s)$ in equation (4.0.2). For starters, it can be shown that the assumption is empirically justified on the daily and weekly time horizon. Figure 4.1 and Figure 4.2 plot the actual final wealth versus the 2-term Taylor series approximation for the final wealth for all 500 of the constituent stocks of the S&P 500 from January 2, 1998 to December 31, 2002. Likewise Figure 4.3 and Figure 4.4 plot the actual growth rate versus the Taylor series approximated growth rate. As is evident, the estimated values depart only very slightly from the two-term empirical Taylor approximation.

More formally, when will the Taylor series approximation be valid? Only when $\mathbf{w}'\mathbf{r}_s$ approaches zero. And how can one ensure that this will take place? For any fixed length of time we consider, we can continue to redefine our time interval for $\mathbf{r}_s$

so that it gets shorter and shorter. Assuming a continuity of prices, we will ensure that $\mathbf{w}'\mathbf{r}_s$ approaches zero and that our Taylor series approximation is valid.

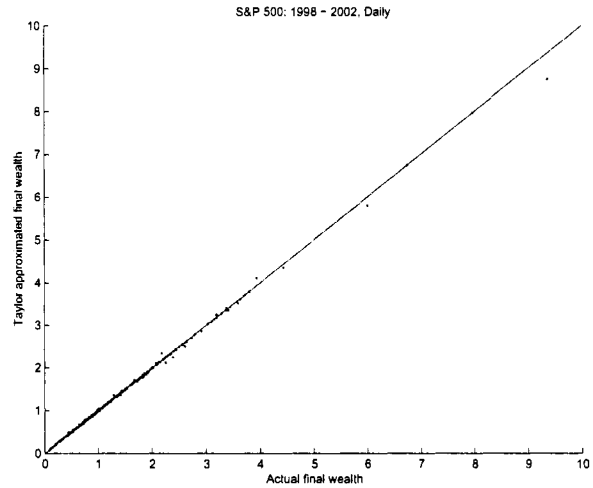

Figure 4.1: A comparison of the actual wealth and Taylor series approximated wealth of S&P 500 stocks from January 2, 1998 to December 31, 2002 using daily data.

4.1 Unconstrained portfolio weights

Proposition 4.1.1. *If we consider all constant rebalanced portfolios in the unconstrained target class* $\tilde{W}$,

$$\tilde{W} = \left\{\mathbf{w} \in \mathbf{R}^k : -\infty < w_i < \infty\right\}$$

instead of Cover's original target class

$$W = \left\{\mathbf{w} \in \mathbf{R}^k : w_i \geq 0, \sum_{i=1}^{k} w_i = 1\right\},$$

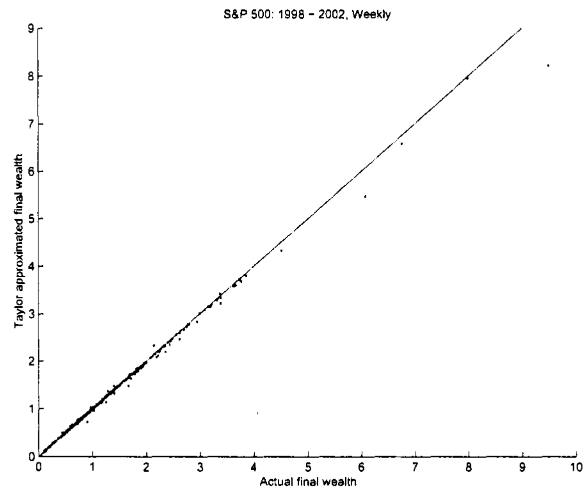

Figure 4.2: A comparison of the actual wealth and Taylor series approximated wealth of S&P 500 stocks from January 2, 1998 to December 31, 2002 using weekly data.

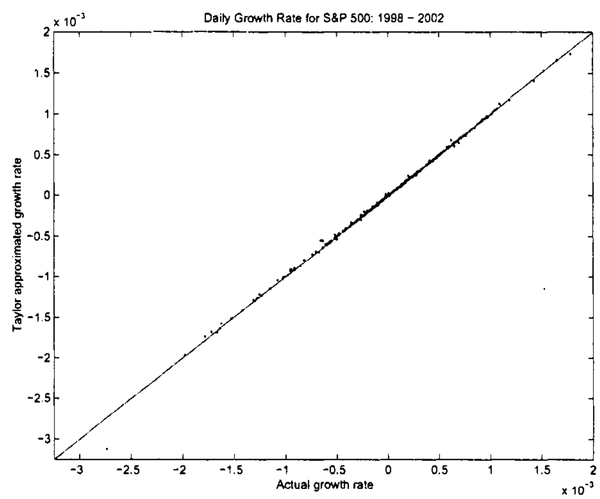

Figure 4.3: A comparison of the actual daily growth rate and the Taylor series approximated daily growth rate of S&P 500 stocks from January 2, 1998 to December 31, 2002.

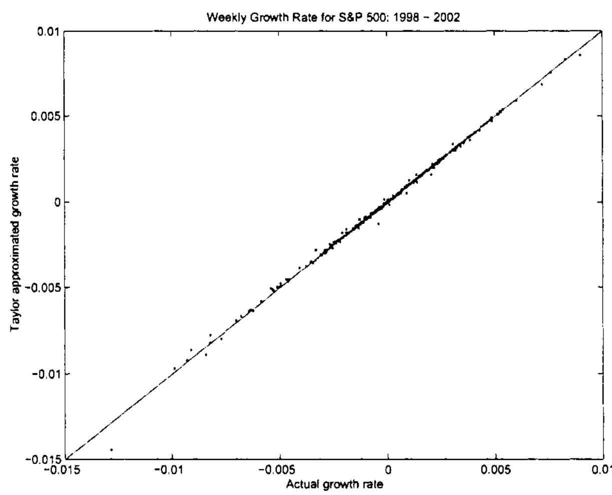

Figure 4.4: A comparison of the actual weekly growth rate and the Taylor series approximated weekly growth rate of S&P 500 stocks from January 2, 1998 to December 31, 2002.

the universal portfolio weights at time $t+1$ are approximately equivalent to the expectation of a multivariate normal random variable with mean $\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t$ and covariance $\frac{1}{t}\bar{\boldsymbol{\Sigma}}_t^{-1}$.

Proof. Substitute equation (4.0.3) for $G_t(\mathbf{w})$ in equation (4.0.1) to give

$$\hat{\mathbf{w}}_{t+1} \approx \tilde{\mathbf{w}}_{t+1} = \frac{\int \mathbf{w} \exp\left\{t\left(\mathbf{w}'\bar{\mathbf{r}}_t - \frac{\mathbf{w}'\bar{\boldsymbol{\Sigma}}_t\mathbf{w}}{2}\right)\right\} d\mathbf{w}}{\int \exp\left\{t\left(\mathbf{w}'\bar{\mathbf{r}}_t - \frac{\mathbf{w}'\bar{\boldsymbol{\Sigma}}_t\mathbf{w}}{2}\right)\right\} d\mathbf{w}}. \tag{4.1.1}$$

By completing the square in equation (4.1.1), we have the more familiar

$$\begin{aligned}
\tilde{\mathbf{w}}_{t+1} &= \frac{\exp\left\{\frac{t\bar{\mathbf{r}}_t'\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t}{2}\right\} \cdot \frac{1}{(2\pi)^{k/2}\left|\frac{\bar{\boldsymbol{\Sigma}}_t^{-1}}{t}\right|^{1/2}} \cdot \int_{\tilde{W}} \mathbf{w} \exp\left\{-\frac{\left(\mathbf{w}-\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t\right)'\left(t\bar{\boldsymbol{\Sigma}}_t\right)\left(\mathbf{w}-\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t\right)}{2}\right\} d\mathbf{w}}{\exp\left\{\frac{t\bar{\mathbf{r}}_t'\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t}{2}\right\} \cdot \frac{1}{(2\pi)^{k/2}\left|\frac{\bar{\boldsymbol{\Sigma}}_t^{-1}}{t}\right|^{1/2}} \cdot \int_{\tilde{W}} \exp\left\{-\frac{\left(\mathbf{w}-\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t\right)'\left(t\bar{\boldsymbol{\Sigma}}_t\right)\left(\mathbf{w}-\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t\right)}{2}\right\} d\mathbf{w}} \\
&= \frac{\frac{1}{(2\pi)^{k/2}\left|\frac{\bar{\boldsymbol{\Sigma}}_t^{-1}}{t}\right|^{1/2}} \cdot \int_{\tilde{W}} \mathbf{w} \exp\left\{-\frac{\left(\mathbf{w}-\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t\right)'\left(t\bar{\boldsymbol{\Sigma}}_t\right)\left(\mathbf{w}-\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t\right)}{2}\right\} d\mathbf{w}}{\frac{1}{(2\pi)^{k/2}\left|\frac{\bar{\boldsymbol{\Sigma}}_t^{-1}}{t}\right|^{1/2}} \cdot \int_{\tilde{W}} \exp\left\{-\frac{\left(\mathbf{w}-\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t\right)'\left(t\bar{\boldsymbol{\Sigma}}_t\right)\left(\mathbf{w}-\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t\right)}{2}\right\} d\mathbf{w}} \\
&= E\left[\mathbf{w}\right]
\end{aligned}$$

where $\mathbf{w} \sim N(\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t, \frac{\bar{\boldsymbol{\Sigma}}_t^{-1}}{t})$. □

This leads to the startling fact that our portfolio weights at time $t+1$ is

$$\hat{\mathbf{w}}_{t+1} \approx \tilde{\mathbf{w}}_{t+1} = \bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t,$$

which is just the well known portfolio weight of a log-optimal investor when the sample mean $\bar{\mathbf{r}}_t$ and sample second moment $\bar{\boldsymbol{\Sigma}}_t$ at time t are used as estimates of the population mean μ and second moment $\boldsymbol{\Sigma}$ at time $t+1$.

4.2 Constrained portfolio weights

With this insight one wonders what would happen if we consider Cover's original target class W? After all, we've just discovered that $\tilde{\mathbf{w}}_{t+1}$ is the expectation of a multivariate normally distributed random variable over a simplex W, scaled by the probability of being in that simplex.

Rewriting equation (4.1.1) as

$$\tilde{\mathbf{w}}_{t+1} = \frac{\int_W f(\mathbf{w})\, e^{-tg(\mathbf{w})} d\mathbf{w}}{\int_W e^{-tg(\mathbf{w})} d\mathbf{w}} \qquad (4.2.1)$$

with

$$f(\mathbf{w}) \stackrel{\text{def}}{=} \mathbf{w}, \qquad g(\mathbf{w}) \stackrel{\text{def}}{=} \frac{\mathbf{w}'\bar{\mathbf{\Sigma}}_t \mathbf{w}}{2} - \mathbf{w}'\bar{\mathbf{r}}_t$$

we can make use of Laplace's method of integration to determine the behavior of $\tilde{\mathbf{w}}_{t+1}$ as $t \to \infty$.

Result 4.2.1. Laplace's Method of Integration. *The leading-order behavior of the integral*

$$I(t) = \int_W f(\mathbf{w}) e^{-tg(\mathbf{w})} d\mathbf{w}$$

can be determined when $f(\mathbf{w})$ is continuous over W, $g(\mathbf{w})$ is twice continuously differentiable over W and $g(\mathbf{w})$ has a minimum over W at $\mathbf{w}^* \in Int\{W\}$. Formally,

$$I(t) = e^{-tg(\mathbf{w}^*)} f(\mathbf{w}^*) \sqrt{\frac{2\pi}{t\,|H_g(\mathbf{w}^*)|}} + O(t^{-1})$$

where $H_g(\mathbf{w}^*)$ represents the Hessian of g evaluated at $\mathbf{w}^*$. The main idea is that for $t >> 1$, the main contribution of the integral comes from a small neighborhood of $\mathbf{w}^*$. See Appendix for an outline of the proof and references.

Proposition 4.2.2. *Cover's Universal Portfolio algorithm is asymptotically equivalent to constrained mean-variance portfolio optimization performed sequentially through time.*

Proof. When Result 4.2.1 is applied to both the numerator and denominator in equation (4.2.1) the resulting approximation has a relative error of order $O(t^{-2})$,

$$\begin{aligned}
\tilde{\mathbf{w}}_{t+1} &= \frac{\int_W f(\mathbf{w})\, e^{-tg(\mathbf{w})} d\mathbf{w}}{\int_W e^{-tg(\mathbf{w})} d\mathbf{w}} \\
&= \frac{e^{-tg(\mathbf{w}^*)} f(\mathbf{w}^*) \sqrt{\frac{2\pi}{t|H_g(\mathbf{w}^*)|}}}{e^{-tg(\mathbf{w}^*)} \sqrt{\frac{2\pi}{t|H_g(\mathbf{w}^*)|}}} + O(t^{-2}) \\
&= f(\mathbf{w}^*) + O(t^{-2}) \\
&= \mathbf{w}^* + O(t^{-2}).
\end{aligned}$$

But what is $\mathbf{w}^*$? According to Result 4.2.1, it is nothing more than the minimum of $g(\mathbf{w}) = \frac{\mathbf{w}'\bar{\boldsymbol{\Sigma}}_t \mathbf{w}}{2} - \mathbf{w}'\bar{\mathbf{r}}_t$ or the maximum of $\mathbf{w}'\bar{\mathbf{r}}_t - \frac{\mathbf{w}'\bar{\boldsymbol{\Sigma}}_t \mathbf{w}}{2}$ over the constrained region W. □

Viewing it this way, one sees that Cover's universal portfolio is nothing more than *constrained* mean-variance optimization performed sequentially through time and the difficulty in computing universal portfolio weights for high dimensions is just the well known difficulty in computing a conditional expectation for the multivariate normal distribution in disguise.

4.3 Approximation technique

Using our two term Taylor expansion for the logarithm in equation (4.2.1), we see that Cover's portfolio weight at time $t+1$ is nothing more than a conditional expectation of a multivariate normal. That is to say,

$$\tilde{\mathbf{w}}_{t+1} \;=\; E\left[\mathbf{w} \middle| \, \mathbf{w} \in W\right]$$

where $\mathbf{w} \sim N(\bar{\boldsymbol{\Sigma}}_t^{-1}\bar{\mathbf{r}}_t, \frac{\bar{\boldsymbol{\Sigma}}_t^{-1}}{t})$ and $W = \left\{ \mathbf{w} \in \mathbf{R}^k : w_i \geq 0, \sum_{i=1}^k w_i = 1 \right\}$. Careful consideration of the properties of a multivariate normal density will give us some insight into the behavior of the universal portfolio weights. For starters, one should consider two cases

1. $\boldsymbol{\Sigma}_t^{-1}\bar{\mathbf{r}}_t \in W$

2. $\boldsymbol{\Sigma}_t^{-1}\bar{\mathbf{r}}_t \notin W$

In the first case, we're blessed with the good fortune of having the mean actually be in the simplex so $\tilde{\mathbf{w}}_{t+1} = \boldsymbol{\Sigma}_t^{-1}\bar{\mathbf{r}}_t$. As for the second case, we can rely upon some empirical facts and properties of the multivariate normal density. The multivariate normal density will be monotonic over our simplex and it turns out that it's quite often very flat. This can be observed by looking at Figure (4.3). In this graph, we took three stocks - Sysco Corp (SYY), Stryker Corp (SYK) and Symantec Corp (SYMC) from January 2, 1998 to December 31, 2002, a period of 1256 trading days.Using the first 625 days to estimate $\boldsymbol{\Sigma}_t^{-1}\bar{\mathbf{r}}_t$ and $\frac{\bar{\boldsymbol{\Sigma}}_t^{-1}}{t}$ we evaluated the multivariate normal density over the simplex and denote its value by the color at that point on the simplex, with

darker colors signifying a larger value. The graph on the right hand side is similar but uses the first 1250 days to estimate parameters. Interestingly enough, we see several things. The multivariate normal density is relatively flat and linear over our simplex. The ratio of the maximum and minimum value over the simplex is less than a factor of three. Hence one potential method of approximating $E[\mathbf{w}|\, \mathbf{w} \in W]$ would be to evaluate the multivariate normal density at all vertices of the simplex and derive the plane that passes through all of the vertices. Then evaluation of $E[\mathbf{w}|\, \mathbf{w} \in W]$ will be analytic and a constant order of computation regardless of the dimension.

4.4 Appendix

Result 4.4.1. Laplace's Method of Integration. *Let $f(\mathbf{w})$ be continuous over W, $g(\mathbf{w})$ be twice continuously differentiable over W and $g(\mathbf{w})$ have a minimum over W at $\mathbf{w}^* \in Int\{W\}$. Then we have*

$$I(t) \;=\; \int_W f(\mathbf{w})e^{-tg(\mathbf{w})}d\mathbf{w} = e^{-tg(\mathbf{w}^*)}f(\mathbf{w}^*)\sqrt{\frac{2\pi}{t\,|H_g(\mathbf{w}^*)|}} + O(t^{-1}).$$

Proof. This is an outline of the proof. See Wong [35], Kass, Tierney and Kadane [22] or Tierney and Kadane [34] for a formal proof of the asymptotic behavior of the Laplace approximation. Rewrite I as

$$\begin{aligned} I(t) &\approx e^{-tg(\mathbf{w}^*)}\int_{\{\mathbf{w}:\|\mathbf{w}-\mathbf{w}^*\|\le\epsilon\}} f(\mathbf{w})e^{-t[g(\mathbf{w})-g(\mathbf{w}^*)]}d\mathbf{w} \\ &\approx e^{-tg(\mathbf{w}^*)}f(\mathbf{w}^*)\int_{\{\mathbf{w}:\|\mathbf{w}-\mathbf{w}^*\|\le\epsilon\}} e^{-t[g(\mathbf{w})-g(\mathbf{w}^*)]}d\mathbf{w} \\ &= e^{-tg(\mathbf{w}^*)}f(\mathbf{w}^*)\int_{\{\mathbf{w}:\|\mathbf{w}-\mathbf{w}^*\|\le\epsilon\}} e^{-t[(\mathbf{w}-\mathbf{w}^*)'\nabla g(\mathbf{w}^*)+\frac{1}{2}(\mathbf{w}-\mathbf{w}^*)'H_g(\mathbf{w}^*)(\mathbf{w}-\mathbf{w}^*)]}d\mathbf{w} \end{aligned}$$

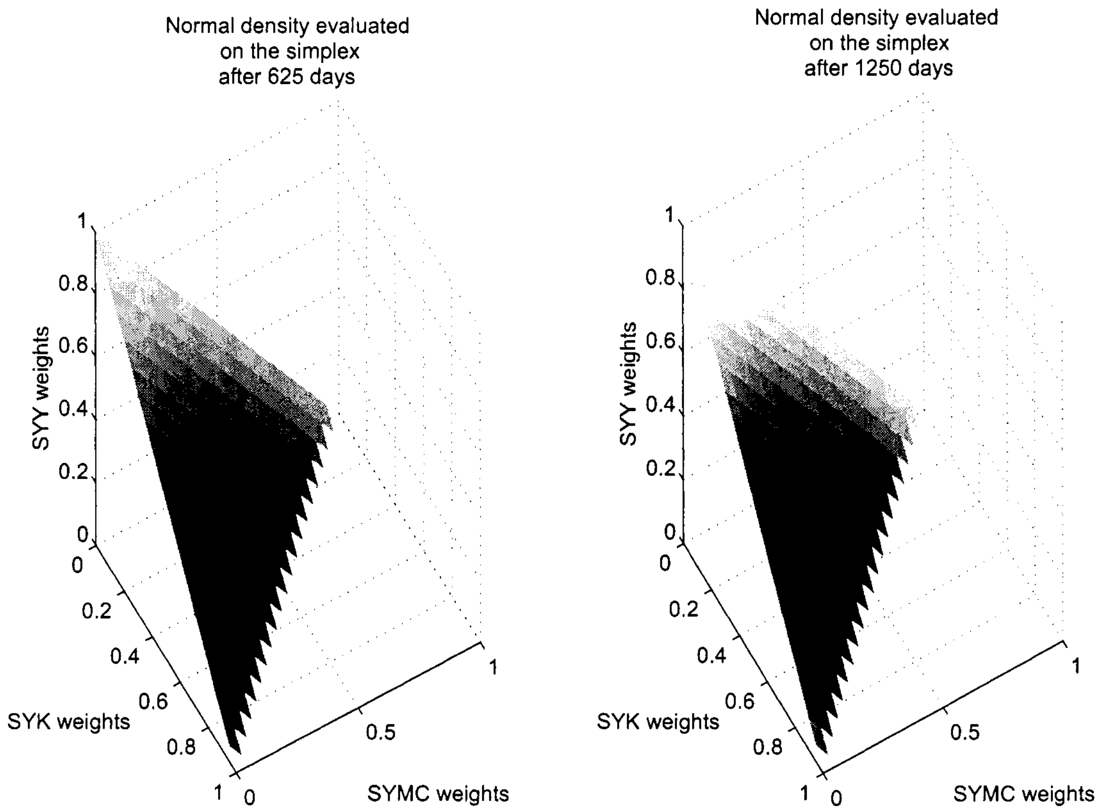

Figure 4.5: Evaluation of the multivariate normal density on the simplex for SYY, SYK and SYMC from January 2, 1998 to December 31, 2002.

$$
\begin{aligned}
&= e^{-tg(\mathbf{w}^*)} f(\mathbf{w}^*) \int_{\{\mathbf{w}:\|\mathbf{w}-\mathbf{w}^*\|\leq\epsilon\}} e^{-\frac{t}{2}(\mathbf{w}-\mathbf{w}^*)' H_g(\mathbf{w}^*)(\mathbf{w}-\mathbf{w}^*)} d\mathbf{w} \\
&= e^{-tg(\mathbf{w}^*)} f(\mathbf{w}^*) \int_{\mathbf{u}} e^{-\frac{t}{2}\mathbf{u}' H_g(\mathbf{w}^*)\mathbf{u}} d\mathbf{u}
\end{aligned}
$$

we see that

$$
I(t) = \int_W f(\mathbf{w}) e^{-tg(\mathbf{w})} d\mathbf{w} \approx e^{-tg(\mathbf{w}^*)} f(\mathbf{w}^*) \sqrt{\frac{2\pi}{t\,|H_g(\mathbf{w}^*)|}}
$$

where $H_g(\mathbf{w}^*)$ represents the Hessian of g evaluated at $\mathbf{w}^*$. The main idea is that for $t >> 1$, the main contribution of the integral comes from a small neighborhood of $\mathbf{w}^*$. □

Chapter 5

Estimation Error

Markowitz's mean-variance framework is one of the most important benchmark models used in practice today. However, his framework requires knowledge of both the mean and covariance of asset returns, which in reality are unknown and have to be estimated from the observed data. Nevertheless, standard practice is to ignore the estimation error and simply treat the sample estimates as the true parameters and plug them back in to get the optimal portfolio weights. The investor will soon see the flaw in this procedure and feel cheated by its ignorance. It turns out, the investor incurs a direct financial cost from this estimation risk.

Let our investor choose $\mathbf{w}$ so as to maximize the Taylor approximated growth rate

$$G(\mathbf{w}) = \mathbf{w}'\mu - \frac{1}{2}\mathbf{w}'\Sigma\mathbf{w}. \tag{5.0.1}$$

As mentioned earlier, when μ and Σ are known, the optimal portfolio weights are

given by

$$\mathbf{w}^* \;=\; \mathbf{\Sigma}^{-1}\mu. \tag{5.0.2}$$

The maximum growth rate that the investor can obtain when the portfolio weights are calculated with the true parameters is

$$\begin{aligned} G(\mathbf{w}^*) &= \mu'\mathbf{\Sigma}^{-1}\mu - \frac{1}{2}\mu'\mathbf{\Sigma}^{-1}\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\mu \\ &= \frac{1}{2}\mu'\mathbf{\Sigma}^{-1}\mu \end{aligned}$$

which is the result of substituting $\mathbf{w}^*$ from equation (5.0.2) in equation (5.0.1). Careful observation leads one to realize that this is a familiar quantity

$$\frac{1}{2}\mu'\mathbf{\Sigma}^{-1}\mu \;=\; \frac{1}{2}\text{ Sharpe Ratio}^2.$$

This means finding the log-optimal portfolio is equivalent to finding the *tangent* portfolio in a Markowitz mean-variance framework. Furthermore, as long as our utility function is within the constant relative risk aversion class, but of differing risk aversion, our optimal growth rate will be a monotone function of the Sharpe ratio.

In practice, μ and $\mathbf{\Sigma}$ are unknown and hence $\mathbf{w}^*$ is not computable. To implement the mean-variance theory of Markowitz, the portfolio weight vectors $\hat{\mathbf{w}}^*$ are typically chosen by a two step process. First, the investor estimates the sample mean and covariance of asset returns up until time t. Second, these sample estimates are used as if they were the true parameters to compute the optimal portfolio weights for use at time $t+1$.

Naturally, the investor who uses $\hat{\mathbf{w}}^*$ should be worse off than the investor who uses $\mathbf{w}^*$, but how do we assess this cost of estimation error? By defining a "loss" function and finding the method that has the lowest "risk", we have a statistical criterion for evaluating various asset allocation procedures.

Definition 5.0.2. Let the "loss" function $L(\mathbf{w}^*, \hat{\mathbf{w}}^*)$ of using $\hat{\mathbf{w}}^*$ instead of $\mathbf{w}^*$ be

$$L(\mathbf{w}^*, \hat{\mathbf{w}}^*) = G(\mathbf{w}^*) - G(\hat{\mathbf{w}}^*).$$

Definition 5.0.3. Define the "risk" function $R(\mathbf{w}^*, \hat{\mathbf{w}}^*)$ to be

$$R(\mathbf{w}^*, \hat{\mathbf{w}}^*) = E\left[L(\mathbf{w}^*, \hat{\mathbf{w}}^*)\right] = E\left[G(\mathbf{w}^*)\right] - E\left[G(\hat{\mathbf{w}}^*)\right].$$

5.1 Data

Shortly, we will make use of 9 different portfolios, each consisting of 20 randomly chosen stocks from the S&P 500 as well as an investment in cash earning the risk free rate. The data for this study was provided by the Center for Research in Security Prices (CRSP) and covers a period of 1256 business days - from January 2, 1998 to December 31, 2002. For comparative purposes, we report the performance of the equal-weighted buy-and-hold portfolio and the equal-weighted constant-rebalanced portfolio for each of the 9 portfolios in Table 5.3.

5.2 Unknown mean μ

Let $\mathbf{\Sigma}$ be known and consider the cost of estimating μ.

PORTFOLIO 1		PORTFOLIO 2		PORTFOLIO 3	
DUK	0.6926	SFA	1.6050	LOW	2.3748
BA	1.0873	PBI	0.7175	APOL	3.8832
GWW	1.2883	APCC	0.8272	DG	0.7944
DNY	0.6539	DG	0.7944	FO	1.7610
XLNX	2.0959	EK	0.5516	BCR	1.5085
NVLS	2.6277	APOL	3.8832	UPC	1.0566
HLT	1.1560	CCU	0.9013	IBM	1.1664
MHP	1.6469	IPG	0.5720	RIG	0.7616
UCL	0.9795	MOLX	1.0518	PHM	2.0897
MXM	0.1876	CTXS	0.8239	DJ	1.0637
HOT	1.2102	MEL	0.9678	SWY	0.5205
STT	1.3171	AIG	1.2483	NTAP	1.7344
AMCC	1.5742	AVP	1.5986	MAS	0.9226
NBR	2.0010	ITW	1.1632	JNJ	1.3655
CMS	0.2628	LSI	1.0373	STJ	3.5904
BMY	0.5150	PG	1.1152	CTB	1.0698
QTRN	0.2460	KMB	1.0939	SLB	0.9555
LTD	1.2271	KEY	0.9778	BSX	1.7177
IPG	0.5720	TXU	0.5237	TJX	1.8298
ADM	0.9130	HMA	1.0464	BNI	0.8470
Fed Funds	1.1322	Fed Funds	1.1322	Fed Funds	1.1322
Buy and Hold	1.1136	Buy and Hold	1.1253	Buy and Hold	1.5307
Equal weight CRB	1.4425	Equal weight CRB	1.5932	Equal weight CRB	1.9610

Table 5.1: Portfolio 1 - 3. The terminal wealth of one dollar invested in 20 randomly chosen stocks from the S&P 500 from January 2, 1998 to December 31, 2002. Buy and hold refers to the portfolio that buys an equal weight on January 2, 1998 and does not rebalance at all. Equal weight CRB refers to the portfolio that rebalances daily to a constant equal weight.

PORTFOLIO 4		PORTFOLIO 5		PORTFOLIO 6	
IR	1.1481	MHP	1.6469	SANM	0.5205
AFL	1.7041	AMGN	2.5717	MU	0.6088
NOC	1.3667	MWD	1.5569	X	0.6416
BUD	1.8635	LPX	0.4340	SBL	0.6201
OXY	1.6169	CR	0.8436	TMO	1.5217
LLY	0.8981	GD	1.6186	EIX	0.4836
HDI	2.5312	ECL	1.7397	MAS	0.9226
AOC	0.5438	APD	1.2508	DRI	2.0393
LTR	1.0852	GLW	0.3214	PAYX	1.5035
SEBL	1.3533	GM	0.8611	WAG	1.2306
JCP	0.5569	ADI	3.1954	PX	1.7085
CSCO	0.9464	BA	1.0873	APD	1.2508
EMR	0.8862	GAS	0.9632	ABK	1.7426
PBI	0.7175	TIF	3.0083	CSX	0.7143
AA	1.2351	GENZ	1.6145	SFA	1.6050
TYC	0.6280	KR	0.6338	NVLS	2.6277
ODP	1.0575	G	0.7339	CTX	1.6373
BAX	1.0098	MMC	1.9210	LLY	0.8981
MCK	0.3153	PVN	0.2453	HSY	1.0123
PCG	0.4685	IPG	0.5720	IBM	1.1664
Fed Funds	1.1322	Fed Funds	1.1322	Fed Funds	1.1322
Buy and Hold	1.0983	Buy and Hold	1.3310	Buy and Hold	1.2185
Equal weight CRB	1.4571	Equal weight CRB	1.6402	Equal weight CRB	1.7543

Table 5.2: Portfolio 4 - 6. The terminal wealth of one dollar invested in 20 randomly chosen stocks from the S&P 500 from January 2, 1998 to December 31, 2002. Buy and hold refers to the portfolio that buys an equal weight on January 2, 1998 and does not rebalance at all. Equal weight CRB refers to the portfolio that rebalances daily to a constant equal weight.

PORTFOLIO 7		PORTFOLIO 8		PORTFOLIO 9	
KSS	2.4871	ADBE	2.7659	MAY	0.6550
GT	0.1509	PSFT	0.8293	TIF	3.0083
SLE	0.8817	T	0.4437	DUK	0.6926
AES	0.1687	BK	0.8395	LU	0.0456
CHIR	1.8684	AA	1.2351	JCI	1.7752
EK	0.5516	PCG	0.4685	MU	0.6088
DTE	1.2807	GE	0.9325	PGR	1.1781
CIEN	1.1041	ROH	1.0416	NTAP	1.7344
DAL	0.2500	BSC	1.9519	MERQ	4.0712
HDI	2.5312	ECL	1.7397	INTU	3.1592
CPB	0.4593	NE	2.3048	TIN	0.9887
COF	1.0278	ETR	1.7689	TE	0.6763
NCC	0.9779	UTX	1.5693	CAG	0.9485
BSX	1.7177	PFE	0.9732	CIN	1.2305
LIZ	2.0059	CA	0.3852	APC	1.1966
RBK	1.9680	SLM	2.6982	TXU	0.5237
AEP	0.6920	MAY	0.6550	SPC	1.1233
TYC	0.6280	FDO	1.7898	GWW	1.2883
SWK	1.4301	RHI	0.8274	XRX	0.1799
HAL	0.5931	TIF	3.0083	DOW	1.1505
Fed Funds	1.1322	Fed Funds	1.1322	Fed Funds	1.1322
Buy and Hold	1.1384	Buy and Hold	1.3981	Buy and Hold	1.3032
Equal weight CRB	1.4309	Equal weight CRB	1.7495	Equal weight CRB	1.6175

Table 5.3: Portfolio 7 - 9. The terminal wealth of one dollar invested in 20 randomly chosen stocks from the S&P 500 from January 2, 1998 to December 31, 2002. Buy and hold refers to the portfolio that buys an equal weight on January 2, 1998 and does not rebalance at all. Equal weight CRB refers to the portfolio that rebalances daily to a constant equal weight.

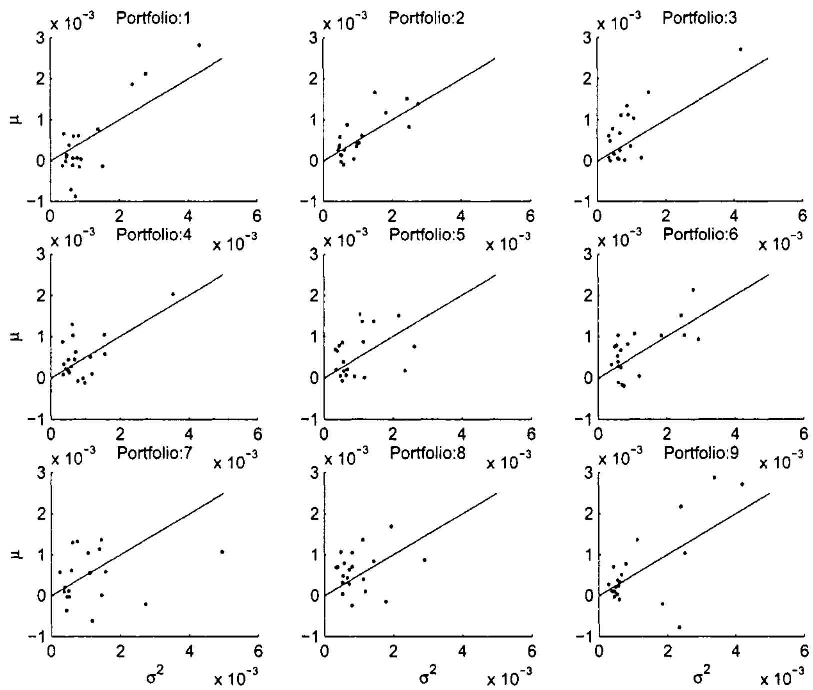

Figure 5.1: μ and σ^2 of return plots for each stock within each of the 9 portfolios.

Result 5.2.1. *By definition, when* $\mathbf{r} \sim N_k(\mu, \mathbf{I}_k)$ *and* $\mu' \neq \mathbf{0}$, $U = \mathbf{r}'\mathbf{r}$ *has a noncentral chi-square distribution* $\chi^2(k, \lambda)$ *with* k *degrees of freedom and noncentrality parameter* $\lambda = \frac{1}{2}\mu'\mu$ *with mean* $E[U] = k + 2\lambda$ *and variance* $\mathrm{Var}[U] = 2(k + 2\lambda)$.

For further details see Ravishanker and Dey [28].

Lemma 5.2.2. *If* $\mathbf{r}_t \sim N_k(\mu, \mathbf{\Sigma})$ *then*

$$\bar{\mathbf{r}}'\mathbf{\Sigma}^{-1}\bar{\mathbf{r}} \sim \frac{1}{t} \cdot \chi^2(k, \frac{t}{2} \cdot \mu'\mathbf{\Sigma}^{-1}\mu),$$

with mean $E[\bar{\mathbf{r}}'\mathbf{\Sigma}^{-1}\bar{\mathbf{r}}] = k/t + \mu'\mathbf{\Sigma}^{-1}\mu$.

Proof. We have $\mathbf{r} \sim N_k(\mu, \mathbf{\Sigma})$ and $\bar{\mathbf{r}} \sim N_k(\mu, \frac{\mathbf{\Sigma}}{t})$. To derive the distribution, we see that

$$\mathbf{r}^* = \sqrt{t}\mathbf{\Sigma}^{-1/2}\bar{\mathbf{r}} \sim N(\sqrt{t}\mathbf{\Sigma}^{-1/2}\mu, \mathbf{I}_k)$$

so using Result 5.2.1, we see that $\mathbf{r}'^*\mathbf{r}^* \sim \chi^2(k, \lambda)$ where $\lambda = \frac{t}{2}\mu'\mathbf{\Sigma}^{-1}\mu$. But this just means

$$\mathbf{r}'^*\mathbf{r}^* = t\bar{\mathbf{r}}'\mathbf{\Sigma}^{-1}\bar{\mathbf{r}} \sim \chi^2(k, \frac{t}{2} \cdot \mu'\mathbf{\Sigma}^{-1}\mu)$$

from which we can see

$$\bar{\mathbf{r}}'\mathbf{\Sigma}^{-1}\bar{\mathbf{r}} \sim \frac{1}{t} \cdot \chi^2(k, \frac{t}{2} \cdot \mu'\mathbf{\Sigma}^{-1}\mu).$$

□

Proposition 5.2.3. *Let* $\mathbf{r}_t \sim N_k(\mu, \mathbf{\Sigma})$, $t = 1, \ldots, T$. *The "risk"* $R(\mathbf{w}^*, \hat{\mathbf{w}}^*)$ *of using* $\hat{\mathbf{w}}^* = \mathbf{\Sigma}^{-1}\bar{\mathbf{r}}$ *instead of* $\mathbf{w}^* = \mathbf{\Sigma}^{-1}\mu$ *is*

$$R(\mathbf{w}^*, \hat{\mathbf{w}}^*) = \frac{k}{2T}.$$

Proof. Substituting, we have

$$
\begin{aligned}
R(\mathbf{w}^*, \hat{\mathbf{w}}^*) &\stackrel{\text{def}}{=} E\left[G(\mathbf{w}^*)\right] - E\left[G(\hat{\mathbf{w}}^*)\right] \\
&= \frac{1}{2}\mu'\boldsymbol{\Sigma}^{-1}\mu - E\left[G(\hat{\mathbf{w}}^*)\right] \\
&= \frac{1}{2}\mu'\boldsymbol{\Sigma}^{-1}\mu - E\left[\hat{\mathbf{w}}'^*\mu - \frac{1}{2}\hat{\mathbf{w}}'^*\boldsymbol{\Sigma}\hat{\mathbf{w}}^*\right] \\
&= \frac{1}{2}\mu'\boldsymbol{\Sigma}^{-1}\mu - \mu'\boldsymbol{\Sigma}^{-1}\mu + \frac{1}{2}E\left[\bar{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{r}}\right] \\
&= \frac{1}{2}\mu'\boldsymbol{\Sigma}^{-1}\mu - \mu'\boldsymbol{\Sigma}^{-1}\mu + \frac{1}{2}\left[\frac{k}{T} + \mu'\boldsymbol{\Sigma}^{-1}\mu\right] \\
&= \frac{k}{2T}.
\end{aligned}
$$

□

Interestingly enough, the expected growth rate with estimation error accounted for, $E\left[G(\hat{\mathbf{w}}^*)\right] = \frac{1}{2}\left[\mu'\boldsymbol{\Sigma}^{-1}\mu - \frac{k}{T}\right]$, need not be positive. That is when $\frac{k}{T} \geq \mu'\boldsymbol{\Sigma}^{-1}\mu$, one would be better advised to not invest at all.

With this observation, one naturally wonders what would happen if the sample estimates of $\mu'\boldsymbol{\Sigma}^{-1}\mu$ for our 9 portfolios of $k = 20$ random stocks from the S&P 500 were used as if they were the truth to evaluate how many years of daily data would be required to ensure a positive expected growth rate once estimation error is accounted for. The results are displayed in Table (5.4) and they are somewhat surprising. For example, consider the fact that we would need anywhere from 5.65 (Portfolio 7) to 16.40 (Portfolio 2) years of daily data in order to ensure $E\left[G(\hat{\mathbf{w}}^*)\right] \geq 0$.

At the same time, let us evaluate the probability of a negative expected growth rate when our investor invests log-optimally but without knowledge of the true parameters.

Portfolio	Time in years	Portfolio	Time in years	Portfolio	Time in years
1	10.65	4	10.64	7	5.65
2	16.40	5	10.56	8	8.80
3	8.73	6	12.66	9	10.25

Table 5.4: The number of years of daily data required to ensure a positive expected growth rate for our 9 portfolios if the sample estimates of $\mu'\boldsymbol{\Sigma}^{-1}\mu$ are used as if they were the truth to evaluate $E\left[G(\hat{\mathbf{w}}^*)\right] = \frac{1}{2}\left[\mu'\boldsymbol{\Sigma}^{-1}\mu - \frac{k}{T}\right]$

Proposition 5.2.4. *The probability of a log-optimal investor incurring a negative growth rate when $\mathbf{r}_t \sim N_k(\mu, \boldsymbol{\Sigma})$ with μ unknown, $\boldsymbol{\Sigma}$ known and k possible assets is*

$$P\left(G(\hat{\mathbf{w}}^*) \leq 0\right) = P\left(\chi_k^2 \geq t \cdot \mu'\boldsymbol{\Sigma}^{-1}\mu\right).$$

Proof. Since $\hat{\mathbf{w}}^* = \boldsymbol{\Sigma}^{-1}\bar{\mathbf{r}}$, we have

$$G(\hat{\mathbf{w}}^*) = \bar{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\mu - \frac{1}{2}\bar{\mathbf{r}}'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{r}},$$

which, after completing the square, can be rewritten as

$$G(\hat{\mathbf{w}}^*) = -\frac{1}{2}\left(\bar{\mathbf{r}} - \mu\right)'\boldsymbol{\Sigma}^{-1}\left(\bar{\mathbf{r}} - \mu\right) + \frac{1}{2}\mu'\boldsymbol{\Sigma}^{-1}\mu.$$

However, since $\bar{\mathbf{r}} \sim N(\mu, t^{-1}\boldsymbol{\Sigma})$, we recognize that

$$G(\hat{\mathbf{w}}^*) \sim \frac{1}{2}\mu'\boldsymbol{\Sigma}^{-1}\mu - \frac{\chi_k^2}{2t}.$$

Hence,

$$P\left(G(\hat{\mathbf{w}}^*) \leq 0\right) = P\left(\frac{1}{2}\mu'\boldsymbol{\Sigma}^{-1}\mu - \frac{\chi_k^2}{2t} \leq 0\right) = P\left(\chi_k^2 \geq t \cdot \mu'\boldsymbol{\Sigma}^{-1}\mu\right).$$

□

In a similar fashion to what we did a moment ago, we can see what this probability would be for our 9 random portfolios of $k = 20$ stocks from the S&P 500 if we, once again, used the sample estimates of $\mu'\boldsymbol{\Sigma}^{-1}\mu$ as if they were the truth. These probabilities are displayed in Table(5.5) and equally as surprising. For example, there is a 60.34% probability that the growth rate for Portfolio 7 is negative.

Portfolio	Probability	Portfolio	Probability	Portfolio	Probability
1	97.74 %	4	97.72 %	7	60.34 %
2	99.87 %	5	97.62 %	8	93.49 %
3	93.20 %	6	99.23 %	9	97.16 %

Table 5.5: Probability of having a negative growth rate for 9 random portfolios of 20 S&P 500 stocks from January 1998 to December 2002.

5.3 Several estimators of Σ

Up until now we have assumed the covariance matrix was known with certainty but unfortunately this is not the case. Some alternatives to using the sample second moment as an estimator are the following:

1. Exponentially smoothed estimate

2. Eigendecomposition

5.3.1 Exponentially smoothed

For a given fixed N, the exponentially smoothed estimator of Σ is

$$\Sigma_t = \frac{1}{1+\alpha+\alpha^2+\cdots+\alpha^N}\left[\mathbf{r_t r_t'}+\alpha\mathbf{r_{t-1}r_{t-1}'}+\cdots+\alpha^N\mathbf{r_{t-N}r_{t-N}'}\right].$$

5.3.2 Eigendecomposition

To define the eigendecomposition estimator of $\mathbf{\Sigma}$, we start off with the decomposition

$$\mathbf{\Sigma}_t = \mathbf{U}_t\mathbf{D}_t\mathbf{U'}_t$$

where $\mathbf{U}_t$ is a matrix of unit length eigenvectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ of $\mathbf{\Sigma}$ arranged in columns

$$\mathbf{U}_t = [\mathbf{X}_1|\mathbf{X}_2|\cdots|\mathbf{X}_n],$$

and $\mathbf{D}_t$ is the diagonal matrix of eigenvalues $d_1, d_2, \ldots, d_n$ of $\mathbf{\Sigma}$,

$$\mathbf{D}_t = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix}.$$

Now with $0 \leq \alpha \leq 1$ we construct the following matrix $\hat{\mathbf{D}}_t$ of "shrunken" eigenvalues as follows

$$\hat{\mathbf{D}}_t = (1-\alpha)\begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix} + \alpha \bar{d}_{\text{grand}} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

where

$$\bar{d}_{\text{grand}} = \frac{1}{n}\sum_{i=1}^{n} d_i.$$

Putting the pieces back together gives us our shrunken covariance estimator

$$\hat{\mathbf{\Sigma}}_t = \mathbf{U}_t \hat{\mathbf{D}}_t \mathbf{U'}_t.$$

5.4 Empirical observation

One empiricical observation made while studying universal portfolios was that the *capitulation portfolio*, the portfolio that invests $1/k$ of its wealth in each of k different assets, frequently outperformed the so called universal portfolio.

Consider the Iroquois Brands Ltd. and Kin Ark Corp example in Cover [9]. These two stocks, listed on the New York Stock Exchange, were studied over a 22 year period from 1963 to 1985 during which time they increased in price by a factor of 8.9 and 4.1 respectively. Cover's universal portfolio would have multiplied a dollar invested

in 1963 by a factor of 37.5, but re-examination of Figure 2 or Table 1 of Cover [9] leads to the startling conclusion that any one who constantly rebalances a portfolio of these two stocks to *any* fraction between 27 and 80 percent did as well or *better* than Cover's universal portfolio.

In fact, consider the 4 pairs of stocks used in Cover [9].

- Iroquois Brands Ltd and Kin Ark Corp.
- Commercial Metals and Kin Ark Corp.
- Commercial Metals and Mei Corp.
- IBM and Coca-cola

In Figure 5.2, we reproduce Figure 8.2, 8.5, 8.6 and 8.7 from Cover [9], which are graphs of the final terminal wealth after 22 years as a function of the constant rebalanced weights. Interestingly enough, the Universal Portfolio never exceeds the naive 50-50 rule, where one maintained equally sized investments in each of the two stocks.

Equally damaging is Table XII of Stoltz and Lugosi [32] which is reproduced here as Table 5.6. For the moment, accept that each column represents one of several online portfolio algorithms which the authors have put in a horse race with one another. For the sake of comparison, they've added a final column "U-CRP" which represents the performance of the *uniform constant rebalanced portfolio* or what we call the *capitulation portfolio.* Interestingly enough, this U-CRP or *capitulation portfolio* is either the best or tied for the best algorithm in 8 of 12 cases.

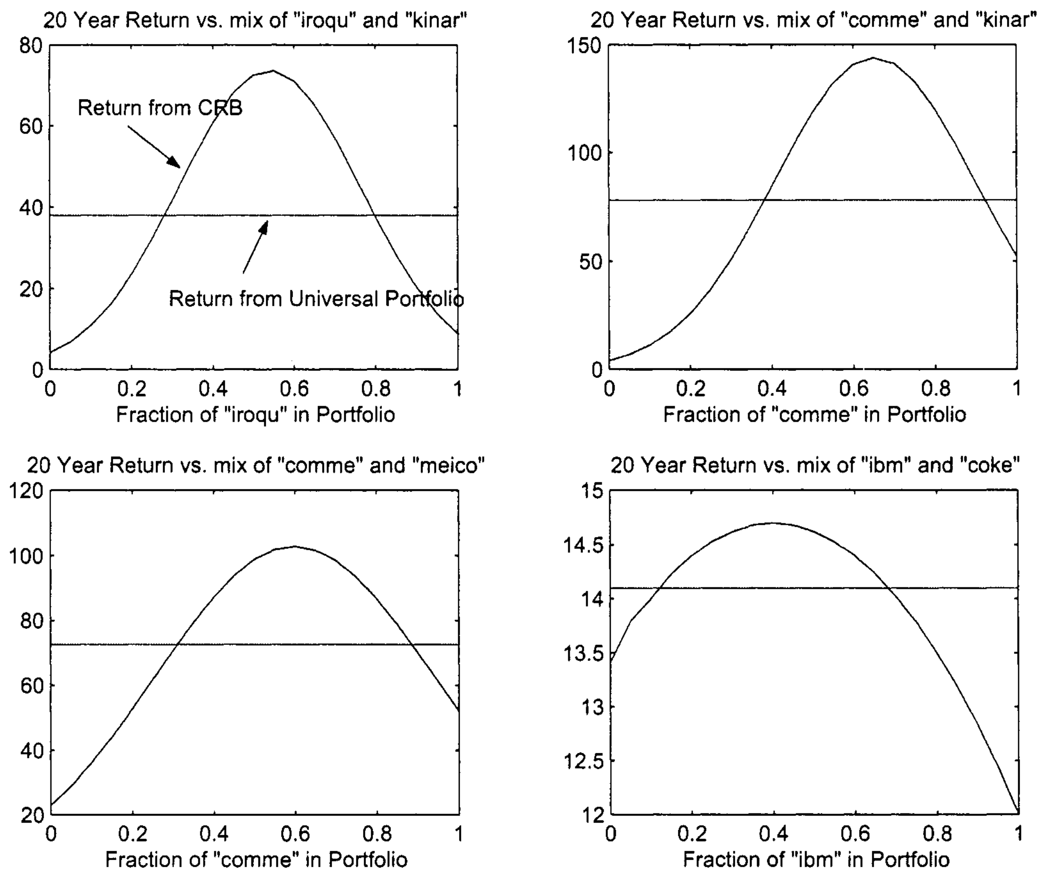

Figure 5.2: Reproduction of Figure 8.2, 8.5, 8.6 and 8.7 from Cover [9]

Portfolio	EG	B1EXP	B1POL	GBH	GBH2	B2POL	Cover's	UBH	U-CRP
L12	10.9	11.1	7.6	10.8	10.1	7.7	11.0	9.4	11.2
M12	17.2	17.1	22.9	17.1	16.9	21.9	17.0	16.7	17.1
H12	36.3	39.0	12.8	34.6	25.3	10.2	37.8	17.6	39.8
L24	13.9	14.0	19.8	14.0	13.5	15.7	14.1	13.1	14.1
H24	26.7	27.8	41.8	27.1	21.8	21.7	27.6	17.2	28.0
A36	20.5	21.2	30.9	20.8	17.5	22.5	20.7	14.5	21.1
L12	12.3	12.4	6.7	12.0	11.1	6.5	12.2	10.1	12.4
M12	16.1	16.2	9.9	15.8	14.8	9.4	16.0	13.9	16.2
H12	78.1	81.0	40.8	67.9	40.2	21.9	76.0	19.5	81.9
L24	14.3	14.4	9.3	14.2	13.1	9.0	14.4	12.0	14.4
H24	38.2	38.7	25.6	38.1	26.1	21.9	38.6	16.7	38.8
A36	26.9	27.1	20.2	27.1	20.2	17.4	27.0	14.5	27.1

Table 5.6: Wealths achieved by the portfolios in Stoltz and Lugosi [5.6]. Monthly rebalancing in the top block. Daily rebalancing in the bottom block

5.5 Shrinkage portfolio

As we previously saw, the naive estimate of $\mathbf{w}^*$ given by $\hat{\mathbf{w}}^* = \mathbf{\Sigma}^{-1}\bar{\mathbf{r}}$ quite often leads to a negative growth rate with high probability. This is a direct consequence of the error in estimating μ, which conceivably one might try to diversify away by the use of another risky portfolio. This is because while both portfolios may have estimation errors, their errors will not be perfectly correlated. But what should this other risky portfolio be?

To answer this, consider the entire financial and statistical spectrum. At the one end we have log-optimal portfolios. They represent the optimistic view that individual companies should have statistically significant and different expected returns and risks from one another. At the other end of the spectrum is the equal weight portfolio. They represent the democratic belief that all companies were created equal, and any one company is statistically indistinguishable from another company in terms of expected returns and risks. Hence a natural candidate for the risky portfolio is to blend the log-optimal portfolio with the equal weight portfolio. The other benefit is that the equal-weight portfolio, is highly correlated with the value weight portfolio, and it is also a mathematically tractable entity for our ensuing calculations.

Hence, let us consider portfolio weights of the form

$$\hat{\mathbf{w}}^* = \hat{\mathbf{w}}^*(\alpha, \beta) = \alpha\mathbf{\Sigma}^{-1}\bar{\mathbf{r}} + \beta\frac{\mathbf{1}}{k} \tag{5.5.1}$$

where α and β are chosen to minimize $R(\mathbf{w}^*, \hat{\mathbf{w}}^*)$.

Proposition 5.5.1. *The portfolio weights which minimize $R(\mathbf{w}^*, \hat{\mathbf{w}}^*)$ are*

$$\hat{\mathbf{w}}^* = \left(\frac{\gamma}{\gamma + \frac{k}{t}}\right) \mathbf{\Sigma}^{-1}\bar{\mathbf{r}} + \left(\frac{\frac{k}{t}}{\gamma + \frac{k}{t}}\right) \frac{\mathbf{1}'\mu}{\mathbf{1}'\mathbf{\Sigma}\mathbf{1}}\,\mathbf{1}$$

where

$$\gamma \stackrel{\text{def}}{=} \mu'\mathbf{\Sigma}^{-1}\mu - \frac{(\mathbf{1}'\mu)^2}{\mathbf{1}'\mathbf{\Sigma}\mathbf{1}}.$$

Proof. Since $E\left[G(\mathbf{w}^*)\right] = \frac{1}{2}\mu'\mathbf{\Sigma}^{-1}\mu$, the risk function

$$R(\mathbf{w}^*, \hat{\mathbf{w}}^*) = E\left[G(\mathbf{w}^*)\right] - E\left[G(\hat{\mathbf{w}}^*(\alpha,\beta))\right] \tag{5.5.2}$$

is minimized when the expected growth rate, $E\left[G(\hat{\mathbf{w}}^*(\alpha,\beta))\right]$, is maximized as a function of α and β.

To this end, substitute the portfolio weights given by equation (5.5.1) in equation (5.0.1) to calculate the expected growth rate as a function of α and β

$$\begin{aligned} E\left[G(\hat{\mathbf{w}}^*(\alpha,\beta))\right] &= E[\hat{\mathbf{w}}^*(\alpha,\beta)]'\mu - \frac{1}{2}E[\hat{\mathbf{w}}^*(\alpha,\beta)'\mathbf{\Sigma}\hat{\mathbf{w}}^*(\alpha,\beta)] \\ &= \left(\beta\frac{\mathbf{1}'}{k}\mu + \alpha\mu'\mathbf{\Sigma}^{-1}\mu\right) - \frac{1}{2}\left(\beta^2\frac{\mathbf{1}'\mathbf{\Sigma}\mathbf{1}}{k^2} + 2\alpha\beta\frac{\mathbf{1}'\mu}{k} + \alpha^2\left(\frac{k}{t} + \mu'\mathbf{\Sigma}^{-1}\mu\right)\right). \end{aligned}$$

Differentiating with respect to α and β, we solve

$$\begin{aligned} \frac{d}{d\alpha} &= \mu'\mathbf{\Sigma}^{-1}\mu - \beta\frac{\mathbf{1}'\mu}{k} - \alpha\left(\frac{k}{t} + \mu'\mathbf{\Sigma}^{-1}\mu\right) = 0 \\ \frac{d}{d\beta} &= \frac{\mathbf{1}'\mu}{k} - \beta\frac{\mathbf{1}'\mathbf{\Sigma}\mathbf{1}}{k^2} - \alpha\frac{\mathbf{1}'\mu}{k} = 0 \end{aligned}$$

to arrive at the optimal choice of α and β that minimize the risk function. These values are

$$\alpha_{\text{optimal}} = \frac{\gamma}{\gamma + \frac{k}{t}}, \qquad \beta_{\text{optimal}} = \left(\frac{\frac{k}{t}}{\gamma + \frac{k}{t}}\right)\frac{k(\mathbf{1}'\mu)}{\mathbf{1}'\mathbf{\Sigma}\mathbf{1}}$$

where $\gamma \stackrel{\text{def}}{=} \mu'\mathbf{\Sigma}^{-1}\mu - \frac{(\mathbf{1}'\mu)^2}{\mathbf{1}'\mathbf{\Sigma}\mathbf{1}}$. $\square$

Chapter 6

Other Universal Portfolios

A large body of literature exists within the machine learning literature presenting alternate *universal portfolios.* Helmbold, Schapire, Singer and Warmuth [18] have suggested one such universal portfolio with respect to the constant rebalanced portfolio target class. The authors seek a new weight vector $\mathbf{w}_{t+1}$ that (approximately) maximizes the following

$$\eta \log(\mathbf{w}_{t+1} \cdot (1 + \mathbf{r}_t)) - d(\mathbf{w}_{t+1}, \mathbf{w}_t)$$

where $\eta > 0$ is the learning rate and d is some measure of distance between the weight vectors that serve as a penalty term. In their paper, the authors employ a Kullback-Leibler divergence for d

$$d(\mathbf{w}_{t+1}, \mathbf{w}_t) = \mathrm{D}_{KL}(\mathbf{w}_{t+1} || \mathbf{w}_t) = \sum_{i=1}^{k} w_{i,t+1} \log \frac{w_{i,t+1}}{w_{i,t}}.$$

By considering the Taylor approximation to this problem, the authors solve the maximization problem via Lagrangian multipliers to end up with a weight update formula

for asset i as

$$w_{i,t+1} = \frac{w_{i,t} \exp\left(\frac{\eta(1+r_{i,t})}{\mathbf{w}_t \cdot (1+\mathbf{r}_t)}\right)}{\sum_{j=1}^{k} w_{j,t} \exp\left(\frac{\eta(1+r_{j,t})}{\mathbf{w}_t \cdot (1+\mathbf{r}_t)}\right)}.$$

The main advantage of this algorithm over Cover's is its computational ease, but the drawback is that it requires absolute bounds on the return process and the procedure fails to come within a polynomial bound of its target wealth X_t^*.

The results of using HSSW's methodology on the 9 portfolios introduced in Chapter 4 for various levels of η are displayed in Table 6.1. Empirically, the authors found "learning rates around $\eta = .05$ are good choices" and moreover "learning rates from 0.01 to 0.15 all achieved great wealth, greater than the wealth achieved by the universal portfolio."

PORTFOLIO	$\eta = .05$	$\eta = .15$	$\eta = .25$	$\eta = .50$
1	1.4361	1.4171	1.3898	1.2973
2	1.5704	1.5235	1.4752	1.3517
3	1.9461	1.9115	1.8697	1.7350
4	1.4389	1.4010	1.3608	1.2519
5	1.6218	1.5856	1.5501	1.4631
6	1.7236	1.6615	1.5990	1.4437
7	1.4206	1.3958	1.3650	1.2669
8	1.7325	1.6984	1.6638	1.5742
9	1.6162	1.6071	1.5885	1.5048

Table 6.1: Terminal wealths of \$ 1 from using Helmbold, Schapire, Singer & Warmuth's algorithm with different smoothing parameter values.

6.1 Following their lead

Taking a cue from the ad-hoc lead of Helmbold, Schapire, Singer and Warmuth (1998) turns out to work well in practice. As just seen, they included a penalty term in their optimization whose sole purpose was to change the weights smoothly over time. In a similar vein, smoothing our weights as follows

$$\mathbf{w}_{t,\text{smooth}} = (1-\alpha)\mathbf{w}_{t,\text{optimal}} + \alpha\mathbf{w}_{t-1,\text{smooth}}$$

leads to the promising results of Table 6.2.

PORTFOLIO	CRB P/L	Log optimal P/L	
		Smoothed $\mathbf{w}$	Raw $\mathbf{w}$
1	1.4425	4.1054	2.4718
2	1.5932	2.0040	1.1051
3	1.9610	3.9668	8.3413
4	1.4571	3.0933	6.0297
5	1.6402	1.3245	1.2117
6	1.7543	1.8495	0.1779
7	1.4309	4.8253	6.1879
8	1.7495	1.7183	0.7043
9	1.6175	2.7024	0.9960

Table 6.2: Terminal wealths investing \$1 from exponentially smoothing the weights ($\alpha = .99$)

As you may have noticed we still have not altered our optimization problem. Instead we have simply exponentially smoothed the weights from our original opti-

mization. An alternative is for us to perform the following optimization

$$\max \mathbf{w_t'}\mu - \frac{1}{2}\mathbf{w_t'}\Sigma\mathbf{w_t} - \eta \left\|\mathbf{w_t} - \mathbf{w_{t-1}}\right\|_1 .$$

Initially this had more appeal than a Kullback-Leibler or L_2 norm since it is not only a smoothness penalty but also a natural penalty for stock "turnover". Unfortunately, this and the Kullback Leibler penalty lead to an unstable optimization problem due to the difficulty in computing the Hessian matrix when the weights are close to zero. Hence, we performed the following optimization

$$\max \mathbf{w_t'}\mu - \frac{1}{2}\mathbf{w_t'}\Sigma\mathbf{w_t} - \eta \left\|\mathbf{w_t} - \mathbf{w_{t-1}}\right\|_2^2$$

with smoothness parameter $\eta = .05$ and $\eta = .10$. Our covariance estimator was calculated from shrinking the eigenvalues of the sample covariance ($\alpha = .20$) and our estimation of μ was the sample mean. Results are displayed in Table 6.3 and Table 6.4 respectively. Log P/L time series for each portfolio when $\eta = .05$ are provided in Figure 6.1.

6.2 Transaction costs

One issue in any portfolio theory is the impact of inevitable friction one must bear by paying transaction costs. For any one given asset in our portfolio at time t we incur the following transaction costs | Number of shares on day t - Number of shares on day $t-1$| $\times$ Transaction cost per share. But if we let $w_{k,t}$ be the weight of asset k

PORTFOLIO	P/L	Ave. Turnover
1	7.2244	.01781
2	1.5692	.01921
3	6.9269	.01741
4	4.4894	.01708
5	2.4233	.01977
6	0.9610	.02482
7	10.3544	.01536
8	1.6261	.02119
9	3.2067	.02274

Table 6.3: Terminal wealths of \$1 from employing an L_2 penalty term ($\eta = .05$)

PORTFOLIO	P/L	Ave. Turnover
1	7.6286	.01283
2	1.9339	.01416
3	8.3407	.01269
4	3.2456	.01387
5	1.7879	.01573
6	1.4416	.01724
7	8.6039	.01188
8	1.6943	.01585
9	4.1245	.01644

Table 6.4: Terminal wealths of \$1 from employing an L_2 penalty term ($\eta = .10$)

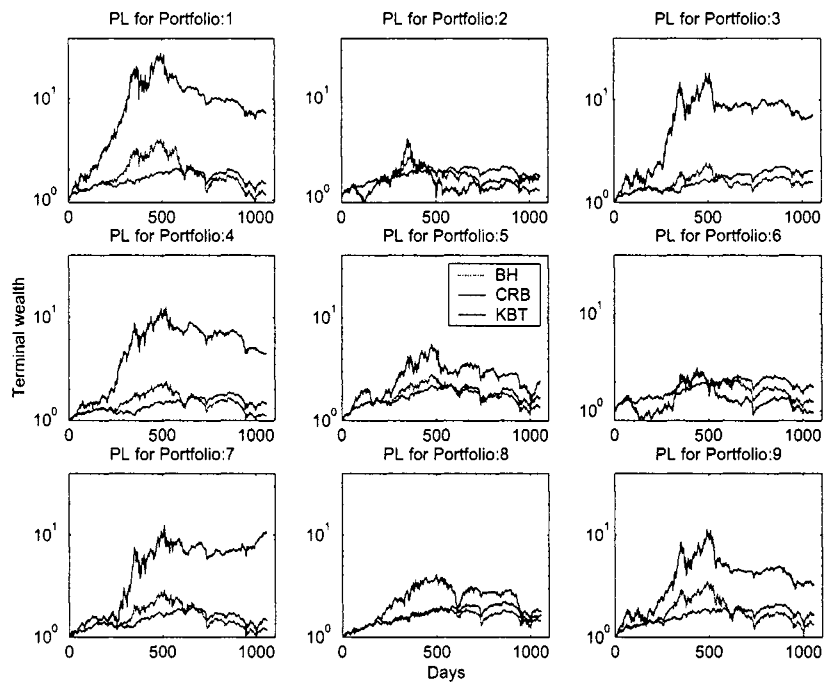

Figure 6.1: Time series of portfolio wealth when $\eta = .05$

on day t, p_t be the portfolio value on day t and $s_{k,t}$ be the price of asset k on day t, we can rewrite the above as

$$\left|\frac{w_{k,t}p_t}{s_{k,t}} - \frac{w_{k,t-1}p_{t-1}}{s_{k,t-1}}\right| \times \text{Transaction cost per share}$$

Moreover, if we allow the transaction cost per share to be a percentage α of the asset price we have

$$\begin{aligned}\text{Transaction cost for stock } k \text{ on day } t &= \left|\frac{w_{k,t}p_t}{s_{k,t}} - \frac{w_{k,t-1}p_{t-1}}{s_{k,t-1}}\right| \times \alpha \times s_{k,t} \\ &= \alpha\left|w_{k,t}p_t - w_{k,t-1}p_{t-1}(1+r_{k,t})\right|\end{aligned}$$

where $r_{k,t}$ is the return of holding asset k from time $t-1$ to time t. If we consider the entire collection of stocks in our portfolio on day t this becomes

$$\alpha\left\| p_t\begin{pmatrix} w_{1,t} \\ w_{2,t} \\ \vdots \\ w_{k,t}\end{pmatrix} - p_{t-1}\begin{pmatrix} 1+r_{1,t} & 0 & 0 & \cdots & 0 \\ 0 & 1+r_{2,t} & 0 & \cdots & 0 \\ 0 & 0 & 1+r_{3,t} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1+r_{k,t}\end{pmatrix}\begin{pmatrix} w_{1,t-1} \\ w_{2,t-1} \\ \vdots \\ w_{k,t-1}\end{pmatrix}\right\|_1 .$$

Now if we sum from $t = 1, \ldots, n$ we have the total transaction costs for our portfolio over the given period of time

$$\text{Total Costs} = \alpha\sum_{t=1}^{n}\left\| p_t\begin{pmatrix} w_{1,t} \\ w_{2,t} \\ \vdots \\ w_{k,t}\end{pmatrix} - p_{t-1}\begin{pmatrix} 1+r_{1,t} & 0 & 0 & \cdots & 0 \\ 0 & 1+r_{2,t} & 0 & \cdots & 0 \\ 0 & 0 & 1+r_{3,t} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1+r_{k,t}\end{pmatrix}\begin{pmatrix} w_{1,t-1} \\ w_{2,t-1} \\ \vdots \\ w_{k,t-1}\end{pmatrix}\right\|_1 .$$

For both daily and weekly rebalancing these costs are outlined in Table (6.5) and Table (6.6).

PORTFOLIO	EIGEN(η = .05,α_{EIGEN}=.20)	COST
1	7.2244	.3978
2	1.5692	.0716
3	6.9269	.2972
4	4.4894	.1976
5	2.4233	.1348
6	0.9610	.0856
7	10.3544	.1984
8	1.6261	.1280
9	3.2067	.2528

Table 6.5: Costs for Daily Rebalancing α_{COST}=.002

PORTFOLIO	EIGEN(η = .10,α_{EIGEN}=.20)	COST
1	8.0443	.3310
2	1.8990	.0498
3	6.3412	.2238
4	3.8960	.1538
5	2.1732	.0872
6	1.0135	.0622
7	10.9347	.1460
8	1.5777	.0916
9	2.8265	.1496

Table 6.6: Costs for Weekly Rebalancing α_{COST}=.002

Chapter 7

Expected Drawdown

One only need spend a short amount of time in the asset management industry to realize that the concept of drawdown is paramount to investors.

Definition 7.0.1. The *current drawdown* of an investment with price process $\{P_t : 0 \leq t \leq T\}$ is defined by

$$D_t = 1 - P_t / \max_{0 \leq s \leq t} P_s.$$

D_t represents the fraction of one's wealth that has been lost since the investment was at its peak. In some ways, it reflects how much regret we have for not having exited our investment at an earlier, more fortuitous time.

7.1 Probabilities and expectations

To gain insight into the properties of drawdown, consider the simplest investment model, geometric Brownian motion. Letting μ denote a constant drift rate and σ^2 denote an instantaneous variance, we assume our investment P_t satisfies the stochastic differential equation

$$dP_t \;=\; \mu P_t dt + \sigma P_t dB_t \tag{7.1.1}$$

where B_t is standard Brownian motion. The solution to this stochastic differential equation is known to be

$$P_t \;=\; P_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma B_t\right\}$$

and without loss of generality we assume our initial investment is \$1, i.e. $P_0 = 1$. Letting

$$\begin{aligned} X_t &= rt + \sigma B_t, \qquad\qquad r \stackrel{\text{def}}{=} \mu - \frac{1}{2}\sigma^2 \\ Y_t &= \max\{X_s : s \in [0,t]\}, \end{aligned}$$

we can rewrite D_t as

$$D_t \;=\; 1 - \frac{e^{X_t}}{e^{Y_t}} = 1 - e^{X_t - Y_t}.$$

Theorem 7.1.1. *Under the model of geometric Brownian motion as given by 7.1.1, the probability of a drawdown at time t exceeding $1-\alpha$ when $0 \le \alpha \le 1$ is*

$$P(D_t \ge 1 - \alpha) \;=\; 1 - \Phi\left(\frac{-\log\alpha + rt}{\sigma\sqrt{t}}\right) - \alpha^{\frac{2r}{\sigma^2}}\left[\Phi\left(\frac{-\log\alpha - rt}{\sigma\sqrt{t}}\right) - 1\right].$$

Proof. From the reflection principle for Brownian motion the joint distribution of (X_t, Y_t) when $y \geq x$ and $y \geq 0$ is known to be

$$P(X_t \leq x, Y_t \leq y) \;=\; \Phi\left(\frac{x - rt}{\sigma\sqrt{t}}\right) - \exp\left\{\frac{2ry}{\sigma^2}\right\}\Phi\left(\frac{x - 2y - rt}{\sigma\sqrt{t}}\right),$$

as one finds, for example, in Harrison ([17], pp 22). Differentiating twice, one has the joint density of (X_t, Y_t) as

$$f_{X_t,Y_t}(x,y) \;=\; \frac{1}{\sqrt{2\pi\sigma^2 t}}\exp\left(\frac{2ry}{\sigma^2} - \frac{(x-2y-rt)^2}{2\sigma^2 t}\right)\left[\frac{2(2y+rt-x)}{\sigma^2 t} - \frac{2r}{\sigma^2}\right]$$

from which we calculate $P(D_t \geq 1-\alpha)$ for $0 \leq \alpha \leq 1$. In particular,

$$\begin{aligned}
P(D_t \geq 1-\alpha) &= P(1 - e^{X_t - Y_t} \geq 1-\alpha) = P(e^{X_t - Y_t} \leq \alpha) \\
&= 1 - P(Y_t - X_t \leq -\log\alpha) \\
&= 1 - \int_0^\infty \int_{y+\log\alpha}^{y} \frac{\exp\left(\frac{2ry}{\sigma^2} - \frac{(x-2y-rt)^2}{2\sigma^2 t}\right)}{\sqrt{2\pi\sigma^2 t}}\left[\frac{2(2y+rt-x)}{\sigma^2 t} - \frac{2r}{\sigma^2}\right] dxdy \\
&= 1 - \Phi\left(\frac{-\log\alpha + rt}{\sigma\sqrt{t}}\right) - \alpha^{\frac{2r}{\sigma^2}}\left[\Phi\left(\frac{-\log\alpha - rt}{\sigma\sqrt{t}}\right) - 1\right].
\end{aligned}$$

□

Theorem 7.1.2. *The expected drawdown at time t is*

$$E(D_t) \;=\; 1 - \frac{2r\Phi\left(\frac{r\sqrt{t}}{\sigma}\right) + 2\exp\left(rt + \frac{\sigma^2 t}{2}\right)(r+\sigma^2)\,\Phi\left(-\frac{\sqrt{t}(r+\sigma^2)}{\sigma}\right)}{2r+\sigma^2} \tag{7.1.2}$$

where $r \stackrel{\text{def}}{=} \mu - \frac{1}{2}\sigma^2$.

Proof. To arrive at the expectation of D_t, observe that $E(D_t) = 1 - E(e^{X_t - Y_t})$ where the density of $e^{X_t - Y_t}$ can be derived from the results of Theorem 7.1.1. That is to say

the density of $e^{X_t - Y_t}$ is

$$\begin{aligned} f(\alpha) &= \frac{d}{d\alpha} P(e^{X_t - Y_t} \leq \alpha) \\ &= \frac{1}{\alpha\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{(-\log\alpha + rt)^2}{2t\sigma^2}\right) + \frac{\alpha^{\frac{2r}{\sigma^2}-1}}{\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{(\log\alpha + rt)^2}{2t\sigma^2}\right) \\ &\quad + \frac{2r}{\sigma^2}\alpha^{\frac{2r}{\sigma^2}-1} - \frac{2r}{\sigma^2}\alpha^{\frac{2r}{\sigma^2}-1}\Phi\left(\frac{-\log\alpha - rt}{\sigma\sqrt{t}}\right), \end{aligned}$$

from which we have

$$\begin{aligned} E(D_t) &= 1 - E(e^{X_t}/e^{Y_t}) = 1 - \int_0^1 \alpha f(\alpha) d\alpha \\ &= 1 - \frac{2r\Phi\left(\frac{r\sqrt{t}}{\sigma}\right) + 2\exp\left(rt + \frac{\sigma^2 t}{2}\right)(r+\sigma^2)\,\Phi\left(-\frac{\sqrt{t}(r+\sigma^2)}{\sigma}\right)}{2r+\sigma^2}. \end{aligned}$$

□

equation (7.1.2) permits an investor to relate μ, σ and t to a drawdown percentage by recalling that r is defined as $\mu - \sigma^2/2$. Likewise, carrying out similar calculations, the variance of D_t can be shown to be

$$\begin{aligned} \mathrm{Var}(D_t) &= \frac{2r\Phi\left(\frac{r\sqrt{t}}{\sigma}\right) + 2\exp\left(2rt + 2\sigma^2 t\right)(r+2\sigma^2)\,\Phi\left(-\frac{\sqrt{t}(r+2\sigma^2)}{\sigma}\right)}{2(r+\sigma^2)} \\ &\quad - \left(\frac{2r\Phi\left(\frac{r\sqrt{t}}{\sigma}\right) + 2\exp\left(rt + \frac{\sigma^2 t}{2}\right)(r+\sigma^2)\,\Phi\left(-\frac{\sqrt{t}(r+\sigma^2)}{\sigma}\right)}{2r+\sigma^2}\right)^2. \end{aligned}$$

7.2 Asymptotics

A natural question after looking at equation (7.1.2) is what happens as we let t approach infinity.

Corollary 7.2.1. *The limiting value of the expected drawdown as t approaches infinity is*

$$\lim_{t\to\infty} E(D_t) = \begin{cases} 1 & \mu \le \sigma^2/2 \\ \frac{\sigma^2}{2r+\sigma^2} = \frac{\sigma^2}{2\mu} & \mu > \sigma^2/2. \end{cases}$$

Proof. Taking care to consider three distinct regions, $r+\sigma^2 \le 0$, $0 \le r+\sigma^2 \le \sigma^2, r > 0$, we can use the following bounds on the tail probability of a Gaussian random variable Z

$$\frac{1}{\sqrt{2\pi}}\frac{1}{z+1/z}\exp\left(-\frac{z^2}{2}\right) \le \Phi(-z) \le \frac{1}{\sqrt{2\pi}}\frac{1}{z}\exp\left(-\frac{z^2}{2}\right), \ z \ge 0$$

to evaluate

$$\begin{aligned} \lim_{t\to\infty} E(D_t) &= 1 - \lim_{t\to\infty}\frac{2r\Phi\left(\frac{r\sqrt{t}}{\sigma}\right) + 2\exp\left(rt+\frac{\sigma^2 t}{2}\right)(r+\sigma^2)\,\Phi\left(-\frac{\sqrt{t}(r+\sigma^2)}{\sigma}\right)}{2r+\sigma^2} \\ &= \begin{cases} 1 & \mu \le \sigma^2/2 \\ \frac{\sigma^2}{2r+\sigma^2} = \frac{\sigma^2}{2\mu} & \mu > \sigma^2/2. \end{cases} \end{aligned}$$

□

Expected drawdown for various levels of σ and t are displayed in Figure (7.1) to exhibit its limiting behavior.

7.3 Controlling drawdown

What are the implications of this for an investor faced with the opportunity to invest in any of k independent investments? In particular, consider k independent investments, each of which follow geometric Brownian motion with constant drift rate μ and

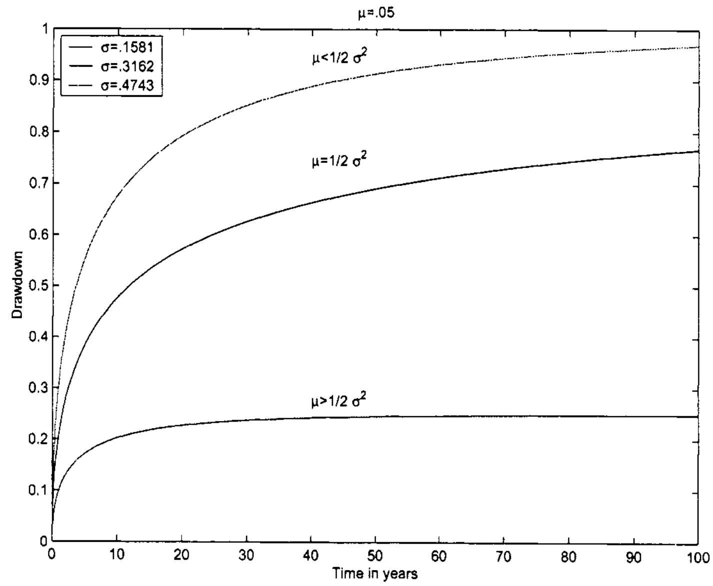

Figure 7.1: Expected drawdown $E(D_t)$ and dependence on σ and t.

instantaneous variance σ^2. Letting $P_{i,t}$ denote an investment in the i^{th} investment, we assume each of our investments $P_{i,t}$ satisfy the stochastic differential equation

$$dP_{i,t} = rP_{i,t}dt + \sigma P_{i,t}dB_{i,t}, \quad i = 1, \ldots, k$$

where $B_{i,t}$ are standard Brownian motions independent of one another.

Within this setup, an investor who decides to buy and hold $1/k^{\text{th}}$ of his wealth in each of the k investments would have the following wealth process at time t

$$W_t = \frac{1}{k}\sum_{i=1}^{k} P_{i,t}.$$

Unfortunately, it is a known fact that the arithmetic average of k geometric Brownian motions is not itself a geometric Brownian motion. Practitioners of continuous time finance have been presented with this very same problem under the guise of

pricing index options. Index options are nothing more than the option to buy or sell an arithmetic weighted portfolio of individual assets at a certain strike price. Their solution is the following – match moments between the true process W_t and a geometric Brownian motion W_t' meant to approximate the true process and solve for the r' and σ'^2 of this new geometric Brownian motion. Matching moments leads to the parameters of our new geometric Brownian motion W_t' being

$$\begin{aligned}
\sigma'^2 &= \log\left(1 - \frac{1}{k} + \frac{e^{\sigma^2}}{k}\right) \\
r' &= \frac{2r + \sigma^2 - \log\left(1 - \frac{1}{k} + \frac{e^{\sigma^2}}{k}\right)}{2}
\end{aligned}$$

To see the expected percentage drawdown from investing in k independent investments, plug these values of r' and σ'^2 back into equation (7.1.2). In other words, $E(D_{k,t})$ equals

$$1 - \frac{2r'\Phi\left(\frac{r'\sqrt{t}}{\sigma'}\right) + 2\exp\left(r't + \frac{\sigma'^2 t}{2}\right)(r' + \sigma'^2)\,\Phi\left(-\frac{\sqrt{t}(r' + \sigma'^2)}{\sigma'}\right)}{2r' + \sigma'^2}. \tag{7.3.1}$$

where $D_{k,t}$ is the drawdown percentage when we buy and hold an equal fraction $1/k$ in each of our k independent investments.

7.4 Simulations

The moment matching method is short on formal justifications, but by simulations one finds that it performs rather well. Table 7.1 gives the average percentage drawdown and the standard deviation of the drawdown from 10,000 simulations with

$r = .0004, \sigma = .01897$ and $t = 2500$ as well as the values of $E(D_{k,2500})$ derived from equation (7.3.1). This choice of r, σ and t is the daily equivalent of 10% annual return and 30% annual volatility over a ten year period. Figure (7.2) displays the histogram of drawdown percentages from this simulation.

The table informs us that the investor with 8 investments as opposed to only 1 investment achieves a nearly 80% reduction in his expected drawdown. Moreover, he achieves a 70% reduction in the standard deviation of this drawdown. So not only does he have lower drawdowns, but he also has less uncertainty as to what these drawdowns will be. The benefits of diversification have cliched, but viewed in this light diversification gains new punch.

k	Simulation $\bar{D}_{k,2500}$	Moment Approximated $E(D_{k,2500})$	k	Simulation $S_{D_{k,2500}}$	Moment Approximated $\sigma_{D_{k,2500}}$
1	.27	.28	1	.21	.21
2	.17	.15	2	.14	.13
4	.10	.07	4	.09	.07
8	.06	.04	8	.06	.04

Table 7.1: Comparing the mean and standard deviations of $D_{k,t}$ from simulation and approximation.

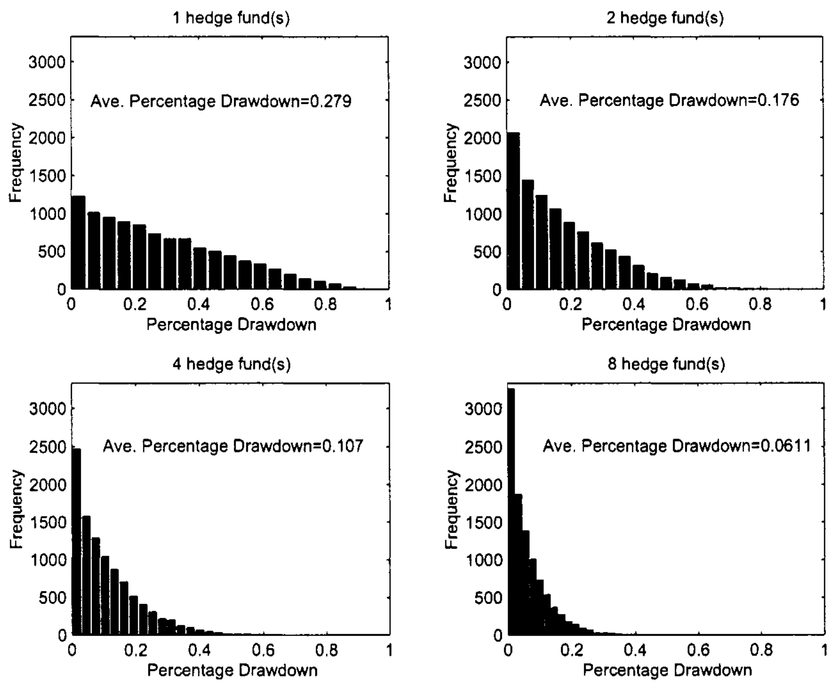

Figure 7.2: Histogram of simulation

Bibliography

[1] Algoet, P. and T.M. Cover, 1988, Asymptotic Optimality and Asymptotic Equipartition Properties of Log-Optimum Investment, *Annals of Probability*, 16, 876-898.

[2] Barron, A. and T.M. Cover, 1988, A Bound on the Financial Value of Information, *IEEE Transactions on Information Theory*, 34, 1097-1100.

[3] Bell, R. and T.M. Cover, 1980, Competitive Optimality of Logarithmic Investment, *Mathematics of Operations Research*, 5, 161-166.

[4] Bell, R. and T.M. Cover, 1988, Game-Theoretic Optimal Portfolios, *Management Science*, 34, 724-733.

[5] Blum, A. and A. Kalai, 1999, Universal Portfolios with and without Transaction Costs, *Machine Learning* 35, 193-205.

[6] Borodin, A., El-Yaniv, R. and V. Gogan, 2000, On the Competitive Theory and Practice of Portfolio Selection.

[7] Breiman, L., 1961, Optimal Gambling Systems for Favorable Games, *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, 1, 65-78.

[8] Cover, T.M., 1984, An Algorithm for Maximizing Expected Log Investment Return, *IEEE Transactions Information Theory*, IT-30, 369-373.

[9] Cover, T.M., 1991, Universal Portfolios, *Mathematical Finance* 1, 1-29.

[10] Cover, T.M. and D. Gluss, 1986, Empirical Bayes Stock Market Portfolios, Advances in Applied Mathematics, 7, 170-181.

[11] Cover, T.M. and E. Ordentlich, 1996, Universal Portfolios with Side Information, *IEEE Transactions on Information Theory* 42, 348-363.

[12] Cover, T.M. and J.A. Thomas, 1991, *Elements of Information Theory*, Wiley, New York.

[13] Foster, D. and R. Vohra, 1999, Regret in the On-line Decision Problem, *Games and Economic Behavior* 29, 7-36.

[14] Frost, P.A. and J.E. Savarino, 1986, An Empirical Bayes Approach to Efficient Portfolio Selection, *Journal of Financial and Quantitative Analysis*, 21, 293-305.

[15] Hakansson, N., 1979, A Characterization of Optimal Multiperiod Portfolio Policies, in *Portfolio Theory, 25 Years After: Essays in Honor of Harry Markowitz*, TIMS Studies in Management Sciences, 11, Amsterdam; North-Holland, 169-177.

[16] Hannan, J.F. and H. Robbins, 1955, Asymptotic Solutions of the COmpound Decision Problem, *Annals of Mathematical Statistics*, 16, 37-51.

[17] Harrison, J.M., *Brownian Motion and Stochastic Flow Systems*, Wiley, New York, 1985.

[18] Helmbold, D.P., Schapire, R.E,, Singer, Y., and M.K. Warmuth, 1998, On-Line Portfolio Selection Using Multiplicative Updates, *Mathematical Finance* 8, 325-347.

[19] Jorion, P., 1986, Bayes-Stein Estimation for Portfolio Analysis, *Journal of Financial and Quantitative Analysis*, 21, 279-292.

[20] Jorion, P., 1991, Bayesian and CAPM Estimators of the Means: Implications for Portfolio Selection, *Journal of Banking and Finance*, 15, 717-727.

[21] Kalai, A. and S. Vempala, 1999, Efficient Algorithms for Universal Portfolios, *Machine Learning* 35, 193-205.

[22] Kass, R.E., Tierney, L. and J.B. Kadane, 1989, Approximate Methods for Assessing Influence and Sensitivity in Bayesian Analysis, *Biometrika*, 76, 663-674.

[23] Kass, R.E. and A.E. Raftery, 1995, Bayes Factors, *Journal of American Statistical Association*, 90, 773-795.

[24] Kelly, J.L., 1956, A New Interpretation of Information Rate, *Bell Systems Technical Journal* 35, 917-926.

[25] Markowitz, H., 1952, Portfolio Selection, *Journal of Finance* 7, 77-91.

[26] Markowitz, H., 1976, Investment for the Long Run: New Evidence for an Old Rule, *Journal of Finance*, 31, 1273-1286.

[27] Merton, R.C. and P.A. Samuelson, 1974, Fallacy of the Log-Normal Approximation to Optimal Portfolio Decision Making over many Periods, *Journal of Financial Economics*, 95, 67-94.

[28] Ravishanker, N. and D.K. Dey, 2002, *A First Course in Linear Model Theory*, Chapman & Hall/CRC, New York.

[29] Sharpe, W.F., 1963, A Simplified Model for Portfolio Analysis, *Management Science*, 9, 277-293.

[30] Steele, J.M., 2000, *Stochastic Calculus and Financial Applications*, Springer-Verlag, New York.

[31] Stein, C., 1956, Inadmissability of the usual estimator for the mean of a multivariate normal distribution, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 197-206.

[32] Stoltz, G. and G. Lugosi, 2003, Internal Regret in on-line Portfolio Selection, Technical Report.

[33] Thorp, E., 1971, Portfolio Choice and the Kelly Criterion, *Proceeding of the Business and Economic Statistics Section, American Statistical Association*, 215-224.

[34] Tierney, L. and J.B. Kadane, 1986, Accurate Approximations for Posterior Moments and Marginal Densities, *Journal of the American Statistical Association*, Vol. 81, 82-86.

[35] Wong, R, 1989, *Asymptotic Approximations of Integrals*, Academic Press, 494-500.